

Guide de l'utilisateur des microdonnées

L'Enquête canadienne sur l'expérience des soins de santé primaires

2007-2008



Statistique
Canada

Statistics
Canada

Canada

Table des matières

1.0	Introduction	5
2.0	Contexte	7
3.0	Objectifs	9
4.0	Concepts et définitions	11
5.0	Méthodologie de l'enquête	13
5.1	Population visée par l'Enquête sur la santé dans les collectivités canadiennes	13
5.2	Plan de sondage de l'Enquête sur la santé dans les collectivités canadiennes	13
5.3	Population visée par l'Enquête canadienne sur l'expérience des soins de santé primaires	14
5.4	Stratégie d'échantillonnage des personnes et taille de l'échantillon	14
6.0	Collecte des données	15
6.1	Conception du questionnaire	16
6.2	Supervision et contrôle de qualité	16
7.0	Traitement des données	17
7.1	Saisie des données	17
7.2	Vérification	17
7.3	Codage des questions ouvertes	17
7.4	Imputation	17
7.5	Création de variables dérivées	19
7.6	Pondération	19
7.7	Suppression de renseignements confidentiels	19
8.0	Qualité des données	21
8.1	Taux de réponse	21
8.2	Erreurs relatives à l'enquête	21
8.2.1	Base de sondage	22
8.2.2	Collecte des données	22
8.2.3	Non-réponse	22
8.2.4	Mesure de l'erreur d'échantillonnage	23
9.0	Lignes directrices pour la totalisation, l'analyse et la diffusion de données	25
9.1	Lignes directrices pour l'arrondissement d'estimations	25
9.2	Lignes directrices pour la pondération de l'échantillon en vue de la totalisation	26
9.3	Définitions de types d'estimations : catégoriques et quantitatives	26
9.3.1	Estimations catégoriques	26
9.3.2	Estimations quantitatives	26
9.3.3	Totalisation d'estimations catégoriques	27
9.3.4	Totalisation d'estimations quantitatives	27
9.4	Lignes directrices pour l'analyse statistique	28
9.5	Lignes directrices pour la diffusion de coefficients de variation	29
9.6	Seuils pour la diffusion des estimations pour l'Enquête canadienne sur l'expérience des soins de santé primaires	31

10.0	Tables de variabilité d'échantillonnage approximative	33
10.1	Comment utiliser les tables de coefficients de variation pour des estimations catégoriques.....	34
10.1.1	Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques	35
10.2	Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance.....	43
10.2.1	Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance	44
10.3	Comment utiliser les tables de coefficients de variation pour effectuer un test t.....	44
10.3.1	Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t	45
10.4	Coefficients de variation pour des estimations quantitatives.....	45
10.5	Tables des coefficients de variation.....	45
10.6	Méthode bootstrap pour estimer la variance	46
10.7	Progiciels statistiques pour estimer la variance.....	46
10.7.1	Autres progiciels.....	47
11.0	Pondération	51
11.1	Procédures de pondération pour l'Enquête sur la santé dans les collectivités canadiennes.....	51
11.2	Procédures de pondération pour l'Enquête canadienne sur l'expérience des soins de santé primaires.....	51
12.0	Questionnaire	55
13.0	Cliché d'enregistrement à valeurs univariées	57

1.0 Introduction

Statistique Canada a mené l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) d'avril à juin 2008 avec l'appui et la collaboration de l'Institut canadien d'information sur la santé et du Conseil canadien de la santé. Ce manuel a été produit pour faciliter la manipulation du fichier de microdonnées portant sur les résultats de l'enquête.

Toutes les questions concernant l'ensemble de données ou son utilisation devraient être adressées à :

Statistique Canada

Services à la clientèle

Division des enquêtes spéciales

Téléphone : 613-951-3321 ou appelez sans frais : 1-800-461-9050

Télécopieur : 613-951-4527

Courriel : des@statcan.ca

2.0 Contexte

Au cours de l'été 2006, le Conseil canadien de la santé (CCS) a communiqué avec la Division des enquêtes spéciales en vue de réaliser le premier cycle de la présente enquête, ce qui a donné lieu à l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) 2006-2007. Le Conseil canadien de la santé a été créé dans le cadre de l'Accord des premiers ministres sur le renouvellement des soins de santé signé en 2003. Il a pour rôle de rendre compte au public des progrès du renouvellement des soins de santé au Canada. L'un de ses objectifs est de fournir, à la population canadienne, une perspective à l'échelle du système sur la réforme des soins de santé, une perspective centrée sur les questions liées à l'évaluation de la responsabilité et de la transparence.

Dès que les résultats de l'enquête de 2006-2007 ont été publiés, le travail a commencé sur le questionnaire 2007-2008. L'Institut canadien d'information sur la santé (ICIS) s'est joint aux membres du CCS et à l'équipe de projet de Statistique Canada afin d'entreprendre l'élaboration de l'enquête de 2007-2008. L'ICIS, qui est devenu commanditaire avec le CCS, est une organisation nationale et indépendante sans but lucratif, qui vise l'amélioration de la santé des Canadiens et du système de soins de santé, en fournissant en temps utile des renseignements fiables et de qualité. L'information de recherche que l'organisation produit porte principalement sur les services de soins de santé, les dépenses de santé et la main-d'œuvre du secteur de la santé, de même que sur des questions touchant la santé de la population.

L'enquête de 2007-2008 différait de celle de 2006-2007 de plusieurs façons. Conjointement aux changements apportés au contenu, particulièrement en ce qui concerne les obstacles à l'accès et à l'utilisation des soins de santé, l'échantillon de l'enquête a été étendu, et une stratégie d'échantillonnage a été élaborée afin de permettre la génération d'estimations nationales et provinciales des résultats de l'enquête. Un nouveau questionnaire a été élaboré et testé en janvier 2008 auprès de groupes de consultation répartis dans quatre villes canadiennes. Le mode de collecte des données a également changé, passant, pour l'enquête de 2006-2007, de l'interview téléphonique au cours de laquelle les réponses sont inscrites sur un formulaire papier à une interview téléphonique assistée par ordinateur (ITAO), pour l'enquête de 2007-2008. La collecte a débuté en avril dans trois bureaux régionaux de Statistique Canada, et elle s'est poursuivie jusqu'à la fin de juin 2008.

3.0 Objectifs

Les objectifs principaux de l'enquête sont de recueillir des données sur les problèmes touchant les Canadiens relativement à leur expérience en matière de soins de santé et de produire des estimations nationales et provinciales. Plus précisément, de dresser le portrait de l'accès aux soins primaires et de leur utilisation ainsi que de fournir de l'information sur les questions qui concernent tout particulièrement les Canadiens atteints d'une maladie chronique et leur expérience du système de soins de santé. Par la suite, les données pourront fournir des renseignements aux fins de l'élaboration de politiques et de stratégies efficaces, tant provinciales que nationales, visant à améliorer les soins de santé pour tous les Canadiens.

Les résultats de l'enquête fourniront une perspective globale de l'expérience des Canadiens en matière de soins de santé et mettront en lumière les problèmes qui touchent les personnes atteintes d'une maladie chronique et permettront d'accroître la sensibilisation à cet égard. Enfin, un des buts ultimes est de fournir des informations qui aideront à faciliter la prise de décisions en ce qui concerne les ressources en fournissant des données de référence qui permettront de suivre l'évolution au fil du temps.

4.0 Concepts et définitions

Étant donné que l'Enquête canadienne sur l'expérience des soins de santé primaires a été réalisée par téléphone, on a fait appel à une terminologie facile à comprendre pour éviter les longues explications des concepts utilisés dans le cadre de l'enquête. L'analyse et l'interprétation des données exigent néanmoins le recours à des concepts et définitions standard. Les questions de l'enquête ont été conçues à la lumière de ces définitions.

Soins de santé primaires : la source principale de soins préventifs, ainsi que les soins continus ou essentiels que les gens reçoivent dans leur collectivité. Ils comprennent les médecins réguliers et les cliniques familiales. Souvent, il s'agit du premier contact du patient avec le système de soins de santé.

Médecins réguliers : tous les médecins ayant un diplôme médical d'une université, payés par le régime d'assurance-maladie provincial. On exclut les médecins qui ne sont pas en titre et ceux qui ne participent pas au régime d'assurance-maladie provincial.

5.0 Méthodologie de l'enquête

L'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) de 2007-2008 a été menée du 14 avril au 30 juin 2008 auprès d'un sous-échantillon des personnes incluses dans l'échantillon de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), cycle 4.1 entre juillet et décembre 2007. Son plan de sondage est donc étroitement lié à celui de l'ESCC. Le plan de l'ESCC 4.1 est décrit brièvement à l'intérieur des sections 5.1 et 5.2¹. Les sections 5.3 et 5.4 décrivent comment l'ECE-SSP s'est écartée du plan de base de l'ESCC 4.1.

5.1 Population visée par l'Enquête sur la santé dans les collectivités canadiennes

Les données de l'ESCC sont recueillies auprès des personnes de 12 ans et plus vivant dans des logements privés des dix provinces et trois territoires. Sont exclus de la base de sondage les habitants des réserves indiennes et des terres de la Couronne, les personnes vivant en institution, les membres à temps plein des Forces armées canadiennes et les habitants de certaines régions éloignées. L'ESCC couvre environ 98 % de la population canadienne âgée de 12 ans et plus.

5.2 Plan de sondage de l'Enquête sur la santé dans les collectivités canadiennes

Afin de produire des estimations fiables pour l'ensemble des 121 régions sociosanitaires (RSS), un échantillon annuel de 65 000 personnes est utilisé. Une stratégie de répartition de l'échantillon, comportant plusieurs étapes, accorde une importance plus ou moins égale aux RSS et aux provinces. L'échantillon est d'abord réparti entre les provinces en fonction de leur taille de population et du nombre de RSS. À l'intérieur des provinces, chaque échantillon est réparti entre les RSS proportionnellement à la racine carrée de la population de la RSS.

L'ESCC utilise trois bases de sondage pour sélectionner les échantillons de ménages : 49 % de l'échantillon des ménages provient de la base aréolaire, 50 % provient d'une base liste de numéros de téléphone et le 1 % restant provient d'une base liste de numéros de téléphone à composition aléatoire (CA). Pour la plupart des RSS, 50 % de l'échantillon est sélectionné à partir d'une base aréolaire et 50 % d'une base liste de numéros de téléphone. Dans deux régions sociosanitaires (Nord-du-Québec et Prairie North), seule la CA est utilisée. Au Nunavut, seule la base aréolaire est utilisée. Au Yukon et dans les Territoires du Nord-Ouest, la plupart de l'échantillon est tiré de la base aréolaire mais un petit échantillon de composition aléatoire est aussi sélectionné dans les capitales.

La base aréolaire conçue pour l'Enquête sur la population active (EPA) du Canada sert de base aréolaire pour l'ESCC. Le plan d'échantillonnage de l'EPA doit donc être considéré dans la sélection de l'échantillon de logements de l'ESCC. Le plan de l'EPA est un plan complexe à deux degrés stratifié dont chaque strate est formée de grappes. L'EPA sélectionne d'abord des grappes par une méthode d'échantillonnage avec probabilité proportionnelle à la taille (PPT) et sélectionne ensuite des logements dans les grappes sélectionnées selon un échantillonnage systématique. L'ESCC a utilisé les grappes de l'EPA qu'elle a stratifié à son tour par RSS. Elle a ensuite sélectionné un échantillon de grappes et de logements dans chaque RSS.

¹ Pour plus de détails sur le plan de sondage de l'ESCC 4.1, voir le Guide pour le fichier de microdonnées à grande diffusion, n° 82M0013GPF au catalogue.

5.3 Population visée par l'Enquête canadienne sur l'expérience des soins de santé primaires

La population visée par l'ECE-SSP 2007-2008 est définie de la même façon que pour l'ESCC 4.1, à l'exception du fait qu'elle se limite aux personnes âgées de 18 ans et plus au 14 avril 2008. Aussi, parce que l'ECE-SSP vise à représenter la population canadienne au début de 2008 mais est sélectionnée à partir des répondants de l'ESCC 4.1, dont la collecte a été réalisée de juillet à décembre 2007, les personnes qui se sont jointes à la population visée entre les deux enquêtes sont exclues. Cette exclusion ne touche pas les gens qui n'avaient pas encore 18 ans au moment de l'ESCC 4.1 puisque celle-ci englobe les personnes âgées de 12 ans et plus.

5.4 Stratégie d'échantillonnage des personnes et taille de l'échantillon

L'ECE-SSP 2007-2008 a été conçue en vue de produire des estimations provinciales et nationales de variables clés sur la santé. La plupart des répondants de l'échantillon de l'ESCC 4.1 sondés entre juillet et décembre 2007 (septembre et décembre pour l'Ontario) ont été utilisés. Presque tout l'échantillon disponible a été utilisé dans 6 des 10 provinces (Terre-Neuve-et-Labrador, Nouveau-Brunswick, Nouvelle-Écosse, Île-du-Prince-Édouard, Manitoba et Saskatchewan) afin de maximiser la proportion estimable minimale (min p) de certaines variables d'intérêt très petites. Autrement, l'enquête a eu recours à une cible de min p de 7 % en vue de déterminer la taille de l'échantillon dans les quatre provinces où un plus grand échantillon était disponible. L'échantillon a été extrait systématiquement de la population âgée de 18 ans et plus de chaque province. La fraction d'échantillonnage est plus petite dans certaines des grandes provinces, particulièrement en Ontario et au Québec. Dans ces provinces, les effets du plan d'échantillonnage sont plus prononcés. Toutefois, la taille plus importante de l'échantillon dans ces régions permet de minimiser les répercussions. Un très petit échantillon de 100 unités a été créé pour les territoires. Ces unités ont été sélectionnées systématiquement et proportionnellement au Yukon, dans les Territoires du Nord-Ouest et au Nunavut en vue de produire uniquement des estimations nationales.

L'échantillon de l'ECE-SSP est composé de 16 482 personnes. Le tableau ci-après indique le nombre de personnes échantillonnées dans chaque province et territoire.

Province et territoire	Taille de l'échantillon
Terre-Neuve-et-Labrador	971
Île-du-Prince-Édouard	671
Nouvelle-Écosse	1 258
Nouveau-Brunswick	1 285
Québec	2 300
Ontario	2 345
Manitoba	1 723
Saskatchewan	1 675
Alberta	2 300
Colombie-Britannique	1 854
Territoires	100
Canada	16 482

6.0 Collecte des données

Une lettre de présentation a été expédiée aux répondants environ une semaine avant le début de la collecte des données de l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) de 2007-2008, qui a eu lieu de la mi-avril jusqu'à la fin de juin 2008, et cette enquête s'est faite par interviews téléphoniques assistées par ordinateur (ITAO).

Le système d'ITAO comporte un certain nombre de modules génériques qui peuvent être rapidement adaptés à la plupart des types d'enquêtes. Un module initial contient un jeu de codes de réponse standard qui traite tous les résultats d'appel possibles, ainsi que les scénarios associés lus par les intervieweurs. Nous avons eu recours à une approche standard pour la présentation de l'organisme et des commanditaires de l'enquête, ainsi que pour la communication du nom et de l'objet de l'enquête, la façon dont les résultats de l'enquête sont utilisés et la durée de l'interview. Nous avons ensuite expliqué aux répondants le processus de sélection des participants à l'enquête, et nous leur avons assuré que leur participation était volontaire et que les renseignements fournis resteraient strictement confidentiels. Des écrans d'aide ont été fournis aux intervieweurs afin d'aider ceux-ci à répondre aux questions fréquemment posées par les répondants.

L'application d'ITAO a fait en sorte que seules les réponses valides soient saisies et que toutes les successions correctes de questions soient respectées. Une fonction de révision a été programmée dans l'application en vue de vérifier la cohérence des réponses, de repérer et corriger les réponses aberrantes, et de déterminer les répondants qui doivent répondre à certaines questions. Cette révision a donné lieu à des données relativement « propres » à la fin du processus de collecte.

Le gestionnaire de l'enquête a rencontré les cadres supérieurs responsables de la collecte pour discuter des problèmes et des questions avant le début de la séance de formation. Le Manuel de l'intervieweur comportait une description du contexte et les objectifs, ainsi qu'une description détaillée des concepts et des définitions propres à l'ECE-SSP 2007-2008. Il renfermait également un glossaire et une série de questions et réponses.

Les intervieweurs ont reçu leur formation sur le contenu de l'enquête dans le cadre d'une séance de formation en classe. De plus, ils ont effectué une série d'interviews simulées pour se familiariser avec l'enquête, ses concepts, ses définitions ainsi que l'application d'ITAO elle-même. La documentation sur les questions et réponses a été fournie aux intervieweurs afin de les aider à répondre aux questions fréquentes des répondants.

La collecte des données a été réalisée par des membres spécialisés du personnel des bureaux de Statistique Canada à Edmonton, Sturgeon Falls et Sherbrooke. La charge de travail et les intervieweurs de chaque bureau étaient gérés par un gestionnaire de projet. L'ordonnanceur utilisé par le système d'ITAO faisait en sorte que les cas soient attribués au hasard aux intervieweurs, et que les répondants soient contactés à différentes heures de la journée et différentes journées de la semaine afin de maximiser les probabilités de contact. Un maximum de 20 appels étaient tentés pour chaque cas déterminé comme étant un numéro de téléphone résidentiel. Dès que le nombre d'appels maximal était atteint, le cas était examiné par un intervieweur principal, qui déterminait s'il fallait effectuer des appels supplémentaires. Un maximum de cinq tentatives d'appel étaient faits pour chaque cas déterminé comme étant un numéro de téléphone inconnu. Si, au cours de ces cinq tentatives d'appel, on constatait qu'un numéro de téléphone appartenait à un ménage, le maximum était élevé à 20 tentatives d'appel.

La durée moyenne d'une interview a été estimée à 22 minutes. Toutefois, la durée des interviews fluctuait selon les circonstances des répondants. Par exemple, la durée moyenne de l'interview était estimée à 30 minutes pour les répondants ayant des états chroniques et à 12 minutes lorsque ce n'était pas le cas.

Il n'y avait aucun retraçage des répondants qui avaient déménagé entre le moment où ils avaient répondu aux questions de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) et celui où ils étaient contactés pour participer à l'ECE-SSP 2007-2008. Toutefois, l'ESCC dispose d'autres coordonnées qui lui permettent de retrouver les répondants, ce qui l'aide beaucoup dans le cas des gens

qui ont déménagé.

6.1 Conception du questionnaire

Le Conseil canadien de la santé (CCS) et l'Institut canadien d'information sur la santé (ICIS) ont fait leur part dans l'élaboration de la version provisoire du questionnaire, notamment en assurant la mise en correspondance avec 27 indicateurs sur la santé élaborés par l'ICIS. Le questionnaire a été remanié afin de tenir compte des objectifs et des indicateurs de recherche des commanditaires. Son contenu a été raccourci de beaucoup, et le déroulement de l'interview a été amélioré. La nouvelle version a été traduite par la Division des langues officielles et de la traduction, puis testée conjointement avec Environics Research Group à l'aide d'interviews réalisées dans les deux langues officielles et dans quatre villes canadiennes. Les tests ont été effectués avec des répondants de divers groupes d'âge et diverses cultures. Une partie du groupe d'essai était composée de gens ayant des troubles chroniques. Les résultats ont permis de modifier davantage le questionnaire. Une fois approuvée la version définitive, des spécifications ont été établies, et une application d'ITAO a été développée puis testée. Des spécifications visant la validité des fourchettes et la cohérence entre les questions ont été incorporées à l'application d'ITAO dans la mesure du possible. Après des essais exhaustifs, l'application a été chargée dans les trois bureaux régionaux de Statistique Canada, où la collecte a débuté le 14 avril 2008.

6.2 Supervision et contrôle de qualité

L'équipe des intervieweurs travaillait sous la supervision des intervieweurs principaux chargés de veiller à faire connaître les concepts et les procédures de l'enquête. L'observation périodique des intervieweurs et l'examen des documents remplis se sont déroulés selon le protocole de collecte.

7.0 Traitement des données

Le principal produit de l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) 2007-2008 est un fichier de microdonnées « épuré ». Ce chapitre présente un bref résumé des phases de traitement inhérentes à la production de ce fichier.

7.1 Saisie des données

Comme la collecte des données s'est faite par interviews téléphoniques assistées par ordinateur, il n'était pas besoin d'avoir un système distinct de saisie des données, car les renseignements étaient entrés directement dans les systèmes des bureaux régionaux par les intervieweurs au moment de l'interview.

7.2 Vérification

La première étape du traitement d'enquête effectué au bureau central a été de remplacer les valeurs « hors limites » incluses dans le fichier de données par des blancs. Ce processus a été conçu pour faciliter les vérifications ultérieures.

Les erreurs dans le déroulement du questionnaire, où l'on a relevé des questions qui ne s'appliquaient pas au répondant (et auxquelles on n'aurait donc pas dû répondre) et qui renfermaient des réponses, constituaient le premier type d'erreurs traitées. Dans ces cas, une vérification par ordinateur a éliminé automatiquement les données superflues en suivant l'ordre du questionnaire dicté par les réponses à des questions antérieures et subséquentes, parfois.

Le second type d'erreurs traitées avait trait à un manque d'information dans les questions pour lesquelles le répondant aurait dû répondre. Pour ce type d'erreur, un code de non-réponse ou « non déclaré » était attribué au poste.

7.3 Codage des questions ouvertes

Il n'y avait aucune question ouverte dans cette enquête.

7.4 Imputation

L'imputation est le processus qui fournit des valeurs valides concernant les variables qui ont été retenues pour être modifiées, soit en raison de renseignements invalides, soit en raison de renseignements manquants. Les nouvelles valeurs sont établies de façon à préserver la structure sous-jacente des données et à garantir que les enregistrements qui en résultent passeront tous les contrôles requis. En d'autres mots, l'objectif n'est pas de reproduire les véritables valeurs des microdonnées mais plutôt d'établir des enregistrements intrinsèquement cohérents qui permettront de produire de bonnes estimations agrégées.

Nous pouvons faire la distinction entre trois types de non-réponse. La non-réponse complète est lorsque le répondant ne fournit pas le nombre minimal de réponses. Ces enregistrements sont supprimés et seront pris en compte lors du processus de pondération (voir le chapitre 11.0). La non-réponse ponctuelle est lorsque le répondant ne fournit pas une réponse à une question mais qu'il passe à la question suivante. Ce type de non-réponse est habituellement traité en utilisant le code « non déclaré » ou en ayant recours à l'imputation. Enfin, la non-réponse partielle est lorsque le répondant fournit le nombre minimal de réponses mais ne termine pas l'interview. Ces enregistrements peuvent être traités comme des cas de non-réponse complète ou ponctuelle.

Puisque les données recueillies sur cette enquêtes portent sur les expériences individuelles des

répondants avec le système de santé, aucune imputation n'a été effectuée.

7.5 Création de variables dérivées

Un certain nombre de données élémentaires incluses dans le fichier de microdonnées ont été calculées en combinant des postes sur le questionnaire pour faciliter l'analyse des données. Ainsi, le caractère rural ou urbain de la communauté (URBRURAL) et les variables des régions métropolitaines de recensement ou des agglomérations de recensement (CMACA) ont été dérivés du code postal.

7.6 Pondération

Le principe qui sous-tend une estimation pour un échantillon probabiliste comme celui de l'ECE-SSP 2007-2008 veut que chacune des personnes incluses dans l'échantillon « représente », en plus d'elle-même, plusieurs autres personnes qui en sont exclues. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne incluse dans l'échantillon représente 50 membres de la population.

La phase de la pondération est une étape où l'on calcule ce nombre (ou poids) pour chaque enregistrement. Ce poids, qui figure dans le fichier de microdonnées, **doit** servir à calculer des estimations significatives à partir de l'enquête. Si, par exemple, le nombre de personnes qui recommanderaient sûrement ou probablement leur fournisseur de soins primaires à un ami ou à un parent, doit être estimé, cette opération s'effectue en sélectionnant les enregistrements se référant aux personnes incluses à l'intérieur de l'échantillon qui présentent cette caractéristique et en additionnant les poids inscrits dans ces enregistrements.

Le chapitre 11.0 renferme des détails au sujet de la méthode utilisée pour calculer ces poids.

7.7 Suppression de renseignements confidentiels

Le fichier partagé contient des données sur tous les répondants qui ont accepté de communiquer leurs données au Conseil canadien de la santé (CCS) et l'Institut canadien d'information sur la santé (ICIS), ainsi que sur ceux qui ont accepté de permettre à Statistique Canada de coupler leurs données d'enquête avec l'Enquête sur la santé dans les collectivités canadiennes (ESCC) 4.1. Il convient de souligner que les données couplées, conformément aux politiques sur la confidentialité de Statistique Canada, ne font pas partie du fichier partagé. Par conséquent, les données couplées ne sont pas communiquées au CCS et ICIS. Comme le taux de partage/couplage était très élevé, soit plus de 94 %, on a jugé que la création d'un fichier maître n'était pas justifiée. Tous les renseignements personnels susceptibles de révéler l'identité des répondants, y compris les noms, numéros de téléphone, adresses municipales et codes postaux, ont été supprimés du fichier partagé.

Il convient de souligner que les fichiers de microdonnées « à grande diffusion » (FMGD) peuvent différer des fichiers « partagés » de l'enquête que conserve Statistique Canada. Ces différences sont habituellement le résultat de mesures prises pour protéger l'anonymat des répondants à une enquête. Les mesures les plus courantes sont la suppression de variables du fichier, le regroupement de valeurs en des catégories plus étendues et le codage de valeurs spécifiques à la catégorie « non déclaré ».

Le fichier partagé de l'enquête comprend quelques renseignements détaillés qui sont inclus sur le FMGD en forme groupé seulement. Ceci comprend

- l'âge précis du répondant;
- le plus haut niveau de scolarité atteint;
- le revenu total du ménage;
- pays de naissance, qui, pour le FMGD a été groupé à « Canada » et « Extérieur du

Canada ».

Pour certaines variables sensibles au risque d'identification d'individus, le FMGD peut avoir subi de la suppression locale, c'est-à-dire que certaines valeurs présentes sur le fichier maître peuvent avoir été codées comme « non déclarées » sur le FMGD. En raison de la petite taille de l'échantillon, tous les enregistrements du Nord ont été exclus du FMGD.

Les utilisateurs ayant besoin d'avoir accès à de l'information exclue des fichiers de microdonnées peuvent acheter des totalisations spéciales. Les estimations produites seront communiquées à l'utilisateur, sous réserve du respect des lignes directrices pour l'analyse et la diffusion dont le chapitre 9.0 de ce document fournit un aperçu.

8.0 Qualité des données

8.1 Taux de réponse

Au total, 16 482 personnes ont été sélectionnées pour participer à l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP). Pour les cas dont on a déterminé qu'ils faisaient ou non partie de l'échantillon (dans le champ ou hors du champ de l'enquête), 127 ne faisaient plus partie de la population cible de l'ECE-SSP (p. ex., en raison d'un décès ou d'un déménagement à l'extérieur du Canada). Des 16 355 personnes estimées admissibles, 11 582 ont répondu à l'enquête et accepté de communiquer leurs données soient transmises aux commanditaires et couplées à leurs réponses de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) 4.1, pour un taux de réponse global de 70,8 %. Le tableau qui suit renferme un résumé des taux de réponse de l'ECE-SSP par province.

Provinces et territoires	Personnes sélectionnées de l'ESCC 4.1	Répondants à l'intérieur du champ de l'enquête	Répondants de l'ECE-SSP	Taux de réponse (%)
Terre-Neuve-et-Labrador	971	961	646	67,2
Île-du-Prince-Édouard	671	662	468	70,7
Nouvelle-Écosse	1 258	1 242	890	71,7
Nouveau-Brunswick	1 285	1 275	846	66,4
Québec	2 300	2 280	1 720	75,4
Ontario	2 345	2 337	1 721	73,6
Manitoba	1 723	1 712	1 059	61,9
Saskatchewan	1 675	1 660	1 200	72,3
Alberta	2 300	2 287	1 676	73,3
Colombie-Britannique	1 854	1 839	1 295	70,4
Territoires	100	100	61	61,0
Canada	16 482	16 355	11 582	70,8

8.2 Erreurs relatives à l'enquête

Les estimations calculées à partir de cette enquête reposent sur un échantillon de personnes. Des estimations légèrement différentes auraient pu être obtenues si un recensement complet avait été effectué en reprenant le même questionnaire et en faisant appel aux mêmes intervieweurs, superviseurs, méthodes de traitement, etc. que ceux effectivement utilisés dans l'enquête. L'écart entre les estimations découlant de l'échantillon et celles que donnerait un dénombrement complet réalisé dans des conditions semblables est appelé erreur d'échantillonnage de l'estimation.

Des erreurs qui ne sont pas liées à l'échantillonnage peuvent se produire à presque toutes les étapes des opérations d'enquête. Les intervieweurs peuvent avoir mal compris les instructions, les enquêtés peuvent se tromper en répondant aux questions, les réponses peuvent être mal saisies sur le questionnaire et des erreurs peuvent survenir lors du traitement et de la totalisation des données. Ces erreurs sont toutes des exemples d'erreurs non dues à l'échantillonnage.

Sur un grand nombre d'observations, les erreurs aléatoires auront peu d'effet sur les estimations calculées à partir de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les

estimations de l'enquête. Énormément de temps et d'efforts ont été consacrés à réduire les erreurs non dues à l'échantillonnage dans l'enquête. Des mesures d'assurance de la qualité ont été prises à chacune des étapes du cycle de collecte et de traitement des données afin de contrôler la qualité des données. Ces mesures comprenaient le recours à des intervieweurs hautement qualifiés, une formation poussée des intervieweurs concernant les procédures d'enquête et du questionnaire, l'observation des intervieweurs en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, des procédures visant à s'assurer que les erreurs de saisie des données étaient réduites au minimum ainsi que des vérifications de la qualité du codage et de contrôle ayant pour but d'attester la logique du traitement.

8.2.1 Base de sondage

Parce que l'ECE-SSP 2007-2008 était une enquête supplémentaire du cycle 4.1 de l'Enquête sur la santé dans les collectivités canadiennes, qui était fondé sur la base de sondage aréolaire, l'Enquête sur la population active (EPA) et la base de sondage téléphonique, y compris la composante de composition aléatoire des numéros de téléphone utilisée par l'ESCC, la qualité des variables d'échantillon dans la base était très élevée et la couverture était très bonne. Il est à noter que la base de sondage de l'ESCC exclut environ 2 % de tous les ménages canadiens. La base de sondage de l'ECE-SSP exclut donc la même proportion de ménages dans les mêmes régions géographiques. Il est peu probable que cette exclusion introduise un biais important dans les données de l'enquête.

Il est à noter que l'entrevue de l'ECE-SSP avait lieu de 4 à 12 mois après celle de l'ESCC 4.1. Pour certaines des personnes choisies à l'ECE-SSP, il n'y avait pas de numéro de téléphone indiqué dans la base de sondage et pour d'autres, le numéro était périmé.

8.2.2 Collecte des données

La formation des intervieweurs pour l'ECE-SSP consistait à lire le Manuel de l'intervieweur et à se familiariser avec le matériel, notamment l'application d'interviews téléphoniques assistées par ordinateur (ITAO). L'information fournie aux intervieweurs comprenait une description du contexte et des objectifs de l'enquête ainsi qu'un glossaire de terminologie et une série de questions et de réponses.

8.2.3 Non-réponse

L'une des principales sources d'erreurs non dues à l'échantillonnage observées dans le cadre des enquêtes est l'effet de la non-réponse sur les résultats de l'enquête. L'étendue d'une non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre simplement à une ou à des questions) à une non-réponse totale. Dans le cas de l'ECE-SSP 2007-2008, il n'y a presque pas eu de non-réponse partielle car une fois le questionnaire amorcé, les répondants avaient tendance à le terminer. S'il y a eu non-réponse totale, c'est parce que l'intervieweur a été incapable de communiquer avec le répondant, ou que le répondant a refusé de participer à l'enquête. Les non-réponses totales ont été traitées en ajustant le poids des personnes qui ont répondu au questionnaire d'enquête de façon à le contrebalancer pour ceux qui n'y ont pas répondu. Voir le chapitre 11.0 pour avoir de plus amples détails sur la correction de la pondération pour la non-réponse. Dans les cas de non-réponse partielle, aucune imputation n'a été effectuée.

8.2.4 Mesure de l'erreur d'échantillonnage

Puisqu'il est inévitable que des estimations établies à partir d'une enquête-échantillon (ou par sondage) soient sujettes à une erreur d'échantillonnage, une saine pratique de la statistique exige que les chercheurs fournissent aux utilisateurs une certaine indication de l'importance de cette erreur d'échantillonnage. Cette section de la documentation renferme un aperçu des mesures de l'erreur d'échantillonnage dont Statistique Canada se sert couramment et que le Bureau conseille vivement aux utilisateurs qui produisent des estimations à partir de ce fichier de microdonnées d'employer également.

La base pour mesurer l'importance potentielle des erreurs d'échantillonnage est l'erreur-type des estimations calculées à partir des résultats d'une enquête.

En raison, cependant, de la diversité des estimations pouvant être produites à partir d'une enquête, l'erreur-type d'une estimation est habituellement exprimée en fonction de l'estimation à laquelle elle se rapporte. La mesure résultante, appelée coefficient de variation (CV) d'une estimation, s'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même et s'exprime en pourcentage de l'estimation.

Par exemple, supposons que d'après les résultats de l'enquête, on estime que 45,1% des Canadiennes ont été diagnostiqués ou traités par un professionnel de la santé pour au moins une des conditions chroniques énumérées dans l'enquête et l'on constate que l'erreur-type de cette estimation est de 0,009. Le coefficient de variation de l'estimation est donc calculé comme suit :

$$\left(\frac{0,009}{0,451} \right) \times 100 \% = 2,0 \%$$

De plus amples renseignements sur le calcul du coefficient de variation, se trouvent au chapitre 10.0.

9.0 Lignes directrices pour la totalisation, l'analyse et la diffusion de données

Ce chapitre de la documentation renferme un aperçu des lignes directrices que doivent respecter les utilisateurs qui totalisent, analysent, publient ou autrement diffusent des données calculées à partir des fichiers de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs de microdonnées de produire les mêmes chiffres que ceux produits par Statistique Canada, tout en étant en mesure d'obtenir des chiffres actuellement inédits de façon conforme à ces lignes directrices établies.

9.1 Lignes directrices pour l'arrondissement d'estimations

Afin que les estimations qui sont destinées à la publication ou à toute autre forme de diffusion qui sont calculées à partir de ces fichiers de microdonnées correspondent à celles produites par Statistique Canada, nous conseillons vivement aux utilisateurs de respecter les lignes directrices qui suivent en ce qui concerne l'arrondissement de telles estimations :

- a) Les estimations dans le corps principal d'un tableau statistique doivent être arrondies à la centaine près à l'aide de la technique d'arrondissement normale. Selon cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1. Par exemple, selon la technique d'arrondissement normale à la centaine près, si les deux derniers chiffres se situent entre 00 et 49, ils sont remplacés par 00 et le chiffre précédent (le chiffre des centaines) reste inchangé. Si les derniers chiffres se situent entre 50 et 99, ils sont remplacés par 00 et le chiffre précédent est augmenté de 1.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes non arrondies correspondantes, puis ensuite être arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement normale.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir de composantes non arrondies (c'est-à-dire des numérateurs et/ou des dénominateurs), puis être arrondis à leur tour à une décimale à l'aide de la technique d'arrondissement normale. Dans le cas d'un arrondissement normal à un seul chiffre, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, le dernier chiffre à conserver est augmenté de 1.
- d) Les sommes et les différences d'agrégats (ou de rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis être arrondies à leur tour à la centaine près (ou à la décimale près) à l'aide de la technique d'arrondissement normale.
- e) Dans les cas, où, en raison de limitations d'ordre technique ou de toutes autres limites, une technique d'arrondissement autre que la technique normale est utilisée produisant des estimations à être publiées ou autrement diffusées différentes des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces différences dans le ou les documents à publier ou à diffuser.
- f) En aucun cas, les utilisateurs ne doivent publier ou autrement diffuser des estimations non arrondies. Des estimations non arrondies laissent entendre qu'elles sont plus précises qu'elles ne le sont en réalité.

9.2 Lignes directrices pour la pondération de l'échantillon en vue de la totalisation

Le plan d'échantillonnage utilisé pour l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) 2007-2008 n'était pas autopondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer le poids d'enquête approprié.

Si les poids appropriés ne sont pas utilisés, les estimations calculées à partir des fichiers de microdonnées ne peuvent être considérées comme représentatives de la population visée par l'enquête et ne correspondront pas à celles produites par Statistique Canada.

Les utilisateurs devraient également prendre note que certains progiciels pourraient peut-être ne pas permettre la production d'estimations correspondant exactement à celles qu'offre Statistique Canada, en raison du mode de traitement du poids par ces progiciels.

9.3 Définitions de types d'estimations : catégoriques et quantitatives

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ECE-SSP 2007-2008, il est utile de décrire les deux principaux types d'estimations ponctuelles des caractéristiques de la population qui peuvent être produites à partir du fichier de microdonnées créé pour l'ECE-SSP.

9.3.1 Estimations catégoriques

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre de personnes qui recommanderaient sûrement ou probablement leur fournisseur de soins primaires à un ami ou à un parent ou la proportion de gens ayant passé au moins une nuit comme patient à l'hôpital, dans une maison de soins infirmiers ou une maison de convalescence au cours des 12 derniers mois constituent des exemples de telles estimations. Une estimation du nombre de personnes possédant une certaine caractéristique peut aussi être désignée comme une estimation d'un agrégat.

Exemples de questions catégoriques :

Q : En général, diriez-vous que votre santé est...?

R : Excellente / Très bonne / Bonne / Passable / Mauvaise

Q : Au cours des 12 derniers mois, avez-vous eu besoin de services de santé de routine ou de suivi?

R : Oui / Non

9.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes et d'autres mesures d'une tendance centrale de quantités reposant sur certains ou sur tous les membres de la population visée par l'enquête. Elles comprennent aussi expressément des estimations de la forme \hat{X} / \hat{Y} où \hat{X} est une estimation de la quantité totale pour la population visée par l'enquête et \hat{Y} , est une estimation du nombre

de personnes dans la population visée par l'enquête qui contribuent à cette quantité totale.

Un exemple d'estimation quantitative est le nombre moyen de nuits passées comme patient à l'hôpital, dans une maison de soins infirmiers ou une maison de convalescence au cours des 12 derniers mois par les personnes ayant passé au moins une nuit dans une telle institution. Le numérateur (\hat{X}) est une estimation du nombre total de nuits passées dans ces institutions dans les 12 derniers mois et son dénominateur (\hat{Y}) est le nombre de personnes ayant déclaré y avoir passé au moins une nuit.

Exemples de questions quantitatives

Q : Combien de nuits dans les 12 derniers mois?

R : |_|_| nuits

Q : En vous incluant, combien de personnes vivent habituellement dans votre ménage?

R : |_|_| personnes

9.3.3 Totalisation d'estimations catégoriques

On peut obtenir des estimations du nombre de gens possédant une certaine caractéristique à partir du fichier de microdonnées en additionnant les poids finals de tous les enregistrements possédant la ou les caractéristiques qui nous intéressent. On obtient des proportions et des rapports de la forme \hat{X} / \hat{Y} en :

- additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le numérateur (\hat{X}),
- additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le dénominateur (\hat{Y}), puis en
- divisant l'estimation a) par celle en b) (\hat{X} / \hat{Y}).

9.3.4 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final de chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Pour obtenir, par exemple, une estimation du nombre total de fois où les femmes ont vu ou consulté un médecin de famille (ou un médecin généraliste) pour des troubles physiques, émotifs ou mentaux au cours des 12 derniers mois, multipliez la valeur déclarée à la question HZ_Q03 (nombre de consultations avec un médecin de famille (ou un médecin généraliste)) par le poids final de l'enregistrement, puis additionnez cette valeur pour tous les enregistrements où la variable SEX = 2 (féminin).

Pour obtenir une moyenne pondérée de la forme \hat{X} / \hat{Y} , le numérateur (\hat{X}) est calculé comme une estimation quantitative et le dénominateur (\hat{Y}) est calculé comme une estimation catégorique. Ainsi, pour estimer le nombre moyen de fois où des femmes ont vu ou consulté un médecin de famille (ou un médecin généraliste) pour des troubles physiques, émotifs ou mentaux au cours des 12 derniers mois,

- a) estimez le nombre total de consultations (\hat{X}) tel qu'il est décrit ci-dessus,
- b) estimez le nombre de femmes (\hat{Y}) incluses dans cette catégorie en additionnant les poids finals de tous les enregistrements où la variable SEX = 2, puis
- c) divisez l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}).

9.4 Lignes directrices pour l'analyse statistique

L'ECE-SSP de 2007-2008 repose sur un plan d'échantillonnage complexe comportant une stratification, de multiples étapes de sélection ainsi que des probabilités inégales de sélection des répondants. L'utilisation des données provenant d'enquêtes aussi complexes présente des problèmes pour les analystes, parce que le plan d'enquête et les probabilités de sélection influent sur les procédures d'estimation et de calcul de la variance qui devraient être utilisées. Il faut utiliser les poids de l'enquête pour que les estimations et les analyses des données de l'enquête soient exemptes de biais.

Bien que de nombreuses procédures d'analyse que l'on trouve à l'intérieur de progiciels statistiques permettent d'utiliser des poids, la signification ou la définition du poids inclus dans ces procédures peut différer de ce qui convient dans le contexte d'une enquête-échantillon, de telle sorte que dans bien des cas, les estimations produites au moyen de ces progiciels sont correctes, mais que les variances calculées sont piètres. Les variances approximatives pour des estimations simples comme des totaux, des proportions et des rapports (pour des variables qualitatives) peuvent être calculées à partir des tables de variabilité d'échantillonnage approximative qui accompagnent les données.

Pour d'autres techniques d'analyse (de régression linéaire, de régression logistique et de l'analyse de variance, par exemple), il existe une méthode qui peut rendre les variances calculées par l'application des progiciels normalisés plus significatives, en intégrant les probabilités inégales de sélection. L'application de cette méthode entraîne une remise à l'échelle des poids de façon à ce que le poids moyen soit de 1.

Supposons, par exemple, qu'il faille effectuer l'analyse de tous les répondants de sexe masculin. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

- 1) sélectionner tous les répondants du fichier qui ont déclaré SEX = homme;
- 2) calculer le poids MOYEN pour ces enregistrements en additionnant les poids originaux des personnes établis à partir du fichier de microdonnées pour ces enregistrements puis diviser cette somme par le nombre de répondants ayant déclaré SEX = homme;
- 3) pour chacun de ces répondants, calculer un poids REMIS À L'ÉCHELLE égal au poids original de la personne divisé par le poids MOYEN;
- 4) effectuer l'analyse portant sur ces répondants en utilisant le poids REMIS À L'ÉCHELLE.

Parce qu'on ne tient toujours compte ni de la stratification ni des grappes du plan d'échantillonnage, les estimations de variance calculées avec cette méthode risquent cependant d'être des sous-estimations.

Il faut connaître les détails du plan d'enquête pour calculer des estimations de variance plus précises. De tels détails ne peuvent être fournis dans le fichier de microdonnées en raisons de la confidentialité. Statistique Canada peut, contre remboursement des frais, calculer des variances qui tiennent compte du plan complet d'échantillonnage pour beaucoup de statistiques.

9.5 Lignes directrices pour la diffusion de coefficients de variation

Avant de diffuser et/ou de publier toutes estimations établies à partir de l'ECE-SSP de 2007-2008, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptable*, *médiocre* et *inacceptable*. Les erreurs d'échantillonnage et non dues à l'échantillonnage, dont il a été question au chapitre 8.0, influencent la qualité des données. Aux fins du présent document, cependant, on ne déterminera le niveau de qualité d'une estimation qu'à partir d'une erreur d'échantillonnage dont rend compte le coefficient de variation indiqué à l'intérieur du tableau qui figure ci-dessous. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8.0 pour être plus pleinement informés des caractéristiques relatives à la qualité de ces données.

On devrait premièrement déterminer le nombre de répondants retenus pour le calcul de l'estimation. Si ce nombre est inférieur à 30, il faudrait considérer l'estimation pondérée comme étant de qualité inacceptable

Pour les estimations pondérées fondées sur les tailles d'échantillon de 30 ou plus, les utilisateurs devraient déterminer le coefficient de variation de l'estimation et suivre les lignes directrices relatives au niveau de qualité qui figurent ci-dessous. Celles-ci devraient être appliquées, pour la détermination du niveau de qualité d'une estimation, aux estimations pondérées arrondies.

On peut considérer qu'il est possible de divulguer toutes les estimations. Celles d'un niveau de qualité médiocre ou inacceptable doivent cependant être accompagnées d'une mise en garde pour avertir les utilisateurs subséquents.

Lignes directrices relatives au niveau de qualité de l'estimation

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	<p>Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent de faibles coefficients de variation, de l'ordre de 0,0 à 16,5 %.</p> <p>Aucune mise en garde n'est requise.</p>
2) Médiocre	<p>Les estimations proviennent d'une taille d'échantillon de 30 ou plus, et présentent des coefficients de variation élevés, de l'ordre de 16,6 à 33,3 %.</p> <p>Ces estimations devraient être signalées par la lettre E (ou un quelconque identificateur similaire). Elles devraient être accompagnées d'une mise en garde avertissant les utilisateurs subséquents des niveaux élevés d'erreur associés aux estimations.</p>
3) Inacceptable	<p>Les estimations proviennent d'une taille d'échantillon inférieure à 30, ou présentent des coefficients de variation très élevés, supérieurs à 33,3 %.</p> <p>Statistique Canada recommande de ne pas diffuser d'estimations de qualité inacceptable. Si un utilisateur choisit cependant de le faire, ces estimations devraient alors être signalées à l'aide de la lettre F (ou d'un quelconque identificateur similaire) et devraient être accompagnées de la mise en garde suivante :</p> <p>« Nous informons l'utilisateur que ces estimations (désignées avec la lettre F) ne respectent pas les normes de qualité de Statistique Canada. Les conclusions qui reposeront sur ces données ne seront pas fiables et seront très probablement invalides. »</p>

9.6 Seuils pour la diffusion des estimations pour l'Enquête canadienne sur l'expérience des soins de santé primaires

Le tableau ci-dessous fournit une indication de la précision des estimations démographiques ainsi que les seuils de diffusion associés aux trois niveaux de qualité de l'estimation présentés à la section précédente. Ces seuils proviennent des tables de coefficients de variation (CV) dont il sera question au chapitre 10.0.

Par exemple, d'après le tableau, la qualité d'une estimation pondérée de 145 000 personnes possédant une caractéristique donnée en Colombie-Britannique est médiocre.

Veillez noter que ces seuils de diffusion correspondent aux estimations de chiffres de population seulement. Dans le cas d'estimations de rapports, les utilisateurs ne devraient pas utiliser la valeur du numérateur (ni le dénominateur) afin de trouver le niveau de qualité de l'estimation correspondant. La règle 4 à la section 10.1 ainsi que l'exemple 4 à la section 10.1.1 expliquent la bonne procédure à suivre dans le cas d'un rapport.

Province / Territoires	CV acceptable 0.0% to 16.5%	CV médiocre 16.6% to 33.3%	CV inacceptable > 33.3%
Terre-Neuve-et-Labrador	37 000 et plus	10 000 à < 37 000	moins de 10 000
Île-du-Prince-Édouard	11 500 et plus	3 000 à < 11 500	moins de 3 000
Nouvelle-Écosse	47 000 et plus	12 000 à < 47 000	moins de 12 000
Nouveau-Brunswick	42 500 et plus	11 000 à < 42 500	moins de 11 000
Québec	328 500 et plus	84 000 à < 328 500	moins de 84 000
Ontario	641 500 et plus	165 500 à < 641 500	moins de 165 500
Manitoba	71 000 et plus	18 500 à < 71 000	moins de 18 500
Saskatchewan	39 500 et plus	10 000 à < 39 500	moins de 10 000
Alberta	135 500 et plus	34 500 à < 135 500	moins de 34 500
Colombie-Britannique	211 000 et plus	54 000 à < 211 000	moins de 54 000
Provinces	394 500 et plus	98 500 à < 394 500	moins de 98 500
Canada	390 500 et plus	97 000 à < 390 500	moins de 97 000

10.0 Tables de variabilité d'échantillonnage approximative

Afin de fournir des coefficients de variation (CV) qui pourraient s'appliquer à une gamme étendue d'estimations catégoriques produites à partir de ce fichier de microdonnées et auxquels il serait facilement possible pour l'utilisateur d'avoir accès, un ensemble de tables de variabilité d'échantillonnage approximative a été produit. Ces tables de CV permettent à l'utilisateur d'obtenir un coefficient de variation approximatif fondé sur la taille de l'estimation calculée à partir des données de l'enquête.

Les coefficients de variation sont calculés à l'aide de la formule de la variance pour un échantillonnage aléatoire simple et en y incorporant un facteur qui reflète la nature du plan d'échantillonnage, qui est à plusieurs degrés et qui prévoit la formation de grappes. Ce facteur, appelé l'effet du plan, a été déterminé en calculant premièrement les effets du plan pour une gamme étendue de caractéristiques, puis en choisissant parmi ceux-ci une valeur modérée (habituellement le 75^e percentile) à utiliser à l'intérieur des tables de CV qui s'appliqueraient ensuite à l'ensemble entier des caractéristiques.

Le tableau ci-dessous indique la valeur modérée des effets du plan, ainsi que les tailles de l'échantillon et les chiffres de population selon la province, qui ont été utilisés pour produire les tables de variabilité d'échantillonnage approximative de l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) 2007-2008.

Province/Territoires	Effet de plan	Taille de l'échantillon	Population
Terre-Neuve-et-Labrador	1,76	646	406 774
Île-du-Prince-Édouard	1,49	468	108 106
Nouvelle-Écosse	1,66	890	732 473
Nouveau-Brunswick	1,79	846	589 261
Québec	2,68	1 720	6 069 167
Ontario	3,22	1 721	9 974 593
Manitoba	2,58	1 059	861 380
Saskatchewan	1,89	1 200	725 057
Alberta	2,46	1 676	2 651 128
Colombie-Britannique	2,28	1 295	3 469 834
Provinces	4,91	11 521	25 587 773
Canada	4,87	11 582	25 661 027

Tous les coefficients de variation inclus dans les tables de variabilité d'échantillonnage approximative sont approximatifs et donc non officiels. Des estimations de la variance réelle pour des variables précises peuvent être obtenues auprès de Statistique Canada, contre remboursement des frais. Étant donné que le CV approximatif est une estimation prudente, l'utilisation de la variance réelle estimée pourrait faire passer l'estimation d'un niveau de qualité à un autre. Par exemple, une estimation *médiocre* pourrait devenir *acceptable* si elle était fondée sur le calcul du CV exact.

Rappelez-vous que : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 30, l'estimation pondérée est très probablement inacceptable et Statistique Canada recommande de ne pas diffuser une telle estimation, quelle que soit la valeur du coefficient de variation.

10.1 Comment utiliser les tables de coefficients de variation pour des estimations catégoriques

Les règles qui suivent devraient permettre à l'utilisateur de déterminer les coefficients de variation approximatifs à partir des tables de variabilité d'échantillonnage approximative pour des estimations du nombre, de la proportion ou du pourcentage de membres de la population visée par l'enquête possédant une certaine caractéristique et pour des rapports et des différences entre de telles estimations.

Règle 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Dans la table de variabilité d'échantillonnage approximative pour la région géographique appropriée, repérez le nombre estimé dans la colonne la plus à gauche (intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) jusqu'au premier chiffre rencontré. Ce chiffre est le coefficient de variation approximatif.

Règle 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou que le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, de personnes prenant des médicaments régulièrement qui avaient subi des effets secondaires au cours des 12 derniers mois est plus fiable que le nombre estimé de personnes prenant des médicaments régulièrement qui avaient subi des effets secondaires au cours des 12 derniers mois. (Remarquez que dans les tables la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

Lorsque la proportion ou que le pourcentage repose sur la population totale de la région géographique visée par la table, le CV de la proportion ou du pourcentage est le même que le CV du numérateur de la proportion ou du pourcentage. Dans ce cas, la règle 1 peut être appliquée.

Lorsque la proportion ou que le pourcentage repose sur un sous-ensemble de la population totale (p. ex. les personnes ayant une maladie chronique), on devrait faire référence à la proportion ou au pourcentage (dans le haut de la table) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche de la table). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

Règle 3 : Estimations de différences entre des agrégats ou des pourcentages

L'erreur-type d'une différence entre deux estimations est approximativement égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. C'est-à-dire que l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1, \hat{X}_2 est l'estimation 2 et α_1 et α_2 sont les coefficients de variation

de \hat{X}_1 et \hat{X}_2 respectivement. Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}}/\hat{d}$. Cette formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais n'est autrement qu'approximative.

Règle 4 : Estimations de rapports

Dans le cas où le numérateur est un sous-ensemble du dénominateur, le rapport devrait être converti en un pourcentage et la règle 2 appliquée. Cela s'appliquerait, par exemple, au cas où le dénominateur est le nombre de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois et le numérateur, le nombre de personnes qui, au cours des 12 derniers mois, avaient éprouvé des difficultés à obtenir ces services dont ils avaient besoin.

Dans le cas où le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois comparativement au nombre de personnes ayant eu besoin de soins immédiats pour un problème de santé mineur au cours de la même période, l'erreur-type du rapport des estimations est approximativement égale à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément, multipliée par \hat{R} . C'est-à-dire que l'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) est :

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement. Le coefficient de variation de \hat{R} est donné par $\sigma_{\hat{R}}/\hat{R}$. La formule tendra à surestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement et à la sous-estimer si \hat{X}_1 et \hat{X}_2 sont corrélés négativement.

Règle 5 : Estimations de différences entre des rapports

Dans ce cas, les règles 3 et 4 sont combinées. On détermine premièrement les CV pour les deux rapports à l'aide de la règle 4, puis on trouve le CV de leur différence au moyen de la règle 3.

10.1.1 Exemples d'utilisation des tables de coefficients de variation pour des estimations catégoriques

Les exemples ci-dessous utilisent des données du fichier de l'ECE-SSP 2007-2008 et sont destinés à aider les utilisateurs à appliquer les règles que nous venons de présenter.

Exemple 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Supposons qu'un utilisateur estime que 14 728 940 le nombre de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA
- 2) L'agrégat estimé 14 728 940 ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 15 000 000.

- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 1,5 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 1,5 %. Le résultat selon lequel il y aurait 14 728 940 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1) de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois précédant l'enquête peut être publié sans réserve.

L'Enquête canadienne sur l'expérience des soins de santé primaires, 2007-2008 - Fichier partagé

Tables de variabilité d'échantillonnage approximative - Canada

NUMÉRATEUR DU PERCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	328,2	326,8	325,1	320,1	311,6	302,8	293,7	284,4	274,8	264,8	254,4	232,2	179,9	103,9
2	232,1	231,1	229,9	226,3	220,3	214,1	207,7	201,1	194,3	187,2	179,9	164,2	127,2	73,4
3	189,5	188,7	187,7	184,8	179,9	174,8	169,6	164,2	158,6	152,9	146,9	134,1	103,9	60,0
4	164,1	163,4	162,6	160,0	155,8	151,4	146,9	142,2	137,4	132,4	127,2	116,1	89,9	51,9
5	146,8	146,1	145,4	143,1	139,3	135,4	131,4	127,2	122,9	118,4	113,8	103,9	80,4	46,4
6	134,0	133,4	132,7	130,7	127,2	123,6	119,9	116,1	112,2	108,1	103,9	94,8	73,4	42,4
7	124,1	123,5	122,9	121,0	117,8	114,4	111,0	107,5	103,9	100,1	96,1	87,8	68,0	39,3
8	116,1	115,5	114,9	113,2	110,2	107,0	103,9	100,6	97,1	93,6	89,9	82,1	63,6	36,7
9	109,4	108,9	108,4	106,7	103,9	100,9	97,9	94,8	91,6	88,3	84,8	77,4	60,0	34,6
10	103,8	103,3	102,8	101,2	98,5	95,7	92,9	89,9	86,9	83,7	80,4	73,4	56,9	32,8
11	99,0	98,5	98,0	96,5	93,9	91,3	88,6	85,8	82,8	79,8	76,7	70,0	54,2	31,3
12	94,8	94,3	93,8	92,4	89,9	87,4	84,8	82,1	79,3	76,4	73,4	67,0	51,9	30,0
13	91,0	90,6	90,2	88,8	86,4	84,0	81,5	78,9	76,2	73,4	70,6	64,4	49,9	28,8
14	87,7	87,3	86,9	85,5	83,3	80,9	78,5	76,0	73,4	70,8	68,0	62,1	48,1	27,8
15	84,8	84,4	83,9	82,6	80,4	78,2	75,8	73,4	70,9	68,4	65,7	60,0	46,4	26,8
16	82,1	81,7	81,3	80,0	77,9	75,7	73,4	71,1	68,7	66,2	63,6	58,1	45,0	26,0
17	79,6	79,3	78,8	77,6	75,6	73,4	71,2	69,0	66,6	64,2	61,7	56,3	43,6	25,2
18	77,4	77,0	76,6	75,4	73,4	71,4	69,2	67,0	64,8	62,4	60,0	54,7	42,4	24,5
...
...
...
750	*****	*****	*****	11,7	11,4	11,1	10,7	10,4	10,0	9,7	9,3	8,5	6,6	3,8
1,000	*****	*****	*****	10,1	9,9	9,6	9,3	9,0	8,7	8,4	8,0	7,3	5,7	3,3
1,500	*****	*****	*****	*****	8,0	7,8	7,6	7,3	7,1	6,8	6,6	6,0	4,6	2,7
2,000	*****	*****	*****	*****	7,0	6,8	6,6	6,4	6,1	5,9	5,7	5,2	4,0	2,3
3,000	*****	*****	*****	*****	*****	5,5	5,4	5,2	5,0	4,8	4,6	4,2	3,3	1,9
4,000	*****	*****	*****	*****	*****	*****	4,6	4,5	4,3	4,2	4,0	3,7	2,8	1,6
5,000	*****	*****	*****	*****	*****	*****	4,2	4,0	3,9	3,7	3,6	3,3	2,5	1,5
6,000	*****	*****	*****	*****	*****	*****	*****	3,7	3,5	3,4	3,3	3,0	2,3	1,3
7,000	*****	*****	*****	*****	*****	*****	*****	*****	3,3	3,2	3,0	2,8	2,1	1,2
8,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	3,0	2,8	2,6	2,0	1,2
9,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,7	2,4	1,9	1,1
10,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,5	2,3	1,8	1,0
12,500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,1	1,6	0,9
15,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,5	0,8
20,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0,7

Note : Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

Exemple 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Supposons qu'un utilisateur estime que $1\,932\,579 / 14\,728\,940 = 13,1\%$ la proportion de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois qui déclaraient avoir eu de la difficulté à obtenir ces services. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Reportez-vous à la table des coefficients de variation pour le CANADA
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois), il faut utiliser à la fois le pourcentage (13,1 %) et la proportion numérateur du pourcentage (1 932 579) pour déterminer le coefficient de variation.
- 3) Le numérateur, 1 932 579, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, soit 2 000 000. De même, l'estimation du pourcentage ne figure dans l'en-tête d'aucune colonne; il faut donc utiliser la proportion qui s'en rapproche le plus, soit 15,0 %.
- 4) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, soit 6,8 %, est le coefficient de variation à employer.
- 5) Le coefficient de variation approximatif de l'estimation est donc 6,8 %. Le résultat selon lequel 13,1 % des personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois déclaraient avoir eu de la difficulté à obtenir ces services peut être publié sans réserve.

Exemple 3 : Estimations de différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime que la proportion de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois qui déclaraient avoir eu de la difficulté à obtenir ces services était de $1\,740\,056 / 13\,939\,092 = 12,5\%$ parmi les personnes ayant déclaré avoir un médecin habituel et de $192\,523 / 771\,022 = 25,0\%$ parmi les personnes n'ayant pas un médecin habituel. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation de la table des coefficients de variation du CANADA de la même façon que celle décrite dans l'exemple 2 donne un CV de 7,8% pour l'estimation pour les personnes ayant un médecin habituel et un CV de 20,1% pour les personnes n'ayant pas un médecin habituel.
- 2) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1 (personnes ayant un médecin habituel), \hat{X}_2 est l'estimation 2 (personnes n'ayant pas un médecin habituel) et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement.

C'est-à-dire que l'erreur-type de la différence $\hat{d} = 0,125 - 0,250 = -0,125$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0,125)(0,078)]^2 + [(0,250)(0,201)]^2} \\ &= \sqrt{(0,000095) + (0,002525)} \\ &= 0,051\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,051 / (-0,125) = -0,408$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 40,8 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas publier cette estimation. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associées à l'estimation.

L'Enquête canadienne sur l'expérience des soins de santé primaires, 2007-2008 - Fichier partagé

Tables de variabilité d'échantillonnage approximative - Canada

NUMÉRATEUR DU PERCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	328,2	326,8	325,1	320,1	311,6	302,8	293,7	284,4	274,8	264,8	254,4	232,2	179,9	103,9
2	232,1	231,1	229,9	226,3	220,3	214,1	207,7	201,1	194,3	187,2	179,9	164,2	127,2	73,4
3	189,5	188,7	187,7	184,8	179,9	174,8	169,6	164,2	158,6	152,9	146,9	134,1	103,9	60,0
4	164,1	163,4	162,6	160,0	155,8	151,4	146,9	142,2	137,4	132,4	127,2	116,1	89,9	51,9
5	146,8	146,1	145,4	143,1	139,3	135,4	131,4	127,2	122,9	118,4	113,8	103,9	80,4	46,4
6	134,0	133,4	132,7	130,7	127,2	123,6	119,9	116,1	112,2	108,1	103,9	94,8	73,4	42,4
7	124,1	123,5	122,9	121,0	117,8	114,4	111,0	107,5	103,9	100,1	96,1	87,8	68,0	39,3
8	116,1	115,5	114,9	113,2	110,2	107,0	103,9	100,6	97,1	93,6	89,9	82,1	63,6	36,7
9	109,4	108,9	108,4	106,7	103,9	100,9	97,9	94,8	91,6	88,3	84,8	77,4	60,0	34,6
10	103,8	103,3	102,8	101,2	98,5	95,7	92,9	89,9	86,9	83,7	80,4	73,4	56,9	32,8
11	99,0	98,5	98,0	96,5	93,9	91,3	88,6	85,8	82,8	79,8	76,7	70,0	54,2	31,3
12	94,8	94,3	93,8	92,4	89,9	87,4	84,8	82,1	79,3	76,4	73,4	67,0	51,9	30,0
13	91,0	90,6	90,2	88,8	86,4	84,0	81,5	78,9	76,2	73,4	70,6	64,4	49,9	28,8
14	87,7	87,3	86,9	85,5	83,3	80,9	78,5	76,0	73,4	70,8	68,0	62,1	48,1	27,8
...
...
100	*****	32,7	32,5	32,0	31,2	30,3	29,4	28,4	27,5	26,5	25,4	23,2	18,0	10,4
125	*****	29,2	29,1	28,6	27,9	27,1	26,3	25,4	24,6	23,7	22,8	20,8	16,1	9,3
150	*****	26,7	26,5	26,1	25,4	24,7	24,0	23,2	22,4	21,6	20,8	19,0	14,7	8,5
200	*****	23,1	23,0	22,6	22,0	21,4	20,8	20,1	19,4	18,7	18,0	16,4	12,7	7,3
250	*****	20,7	20,6	20,2	19,7	19,1	18,6	18,0	17,4	16,7	16,1	14,7	11,4	6,6
300	*****	*****	18,8	18,5	18,0	17,5	17,0	16,4	15,9	15,3	14,7	13,4	10,4	6,0
350	*****	*****	17,4	17,1	16,7	16,2	15,7	15,2	14,7	14,2	13,6	12,4	9,6	5,6
400	*****	*****	16,3	16,0	15,6	15,1	14,7	14,2	13,7	13,2	12,7	11,6	9,0	5,2
450	*****	*****	15,3	15,1	14,7	14,3	13,8	13,4	13,0	12,5	12,0	10,9	8,5	4,9
500	*****	*****	14,5	14,3	13,9	13,5	13,1	12,7	12,3	11,8	11,4	10,4	8,0	4,6
750	*****	*****	*****	11,7	11,4	11,1	10,7	10,4	10,0	9,7	9,3	8,5	6,6	3,8
1,000	*****	*****	*****	10,1	9,9	9,6	9,3	9,0	8,7	8,4	8,0	7,3	5,7	3,3
1,500	*****	*****	*****	*****	8,0	7,8	7,6	7,3	7,1	6,8	6,6	6,0	4,6	2,7
2,000	*****	*****	*****	*****	7,0	6,8	6,6	6,4	6,1	5,9	5,7	5,2	4,0	2,3
3,000	*****	*****	*****	*****	*****	5,5	5,4	5,2	5,0	4,8	4,6	4,2	3,3	1,9
4,000	*****	*****	*****	*****	*****	*****	4,6	4,5	4,3	4,2	4,0	3,7	2,8	1,6
5,000	*****	*****	*****	*****	*****	*****	4,2	4,0	3,9	3,7	3,6	3,3	2,5	1,5
6,000	*****	*****	*****	*****	*****	*****	*****	3,7	3,5	3,4	3,3	3,0	2,3	1,3
7,000	*****	*****	*****	*****	*****	*****	*****	*****	3,3	3,2	3,0	2,8	2,1	1,2
8,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	3,0	2,8	2,6	2,0	1,2
9,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,7	2,4	1,9	1,1
10,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,5	2,3	1,8
12,500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2,1	1,6
15,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1,5
20,000	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0,8
														0,7

NOTE: for correct usage of these tables please refer to microdata documentation.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime qu'au cours des 12 derniers mois, 14 728 940 de personnes avaient eu besoin de services de santé de routine ou de suivi, comparativement à 6 676 981 de personnes qui avaient eu besoin de soins immédiats nécessaires pour un problème de santé mineur. L'utilisateur est intéressé à comparer ces deux estimations sous la forme d'un rapport. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre de personnes qui avaient eu besoin de services de santé de routine ou de suivi. Le dénominateur de l'estimation (\hat{X}_2) est le nombre de personnes qui avaient eu besoin de soins immédiats nécessaires pour un problème de santé mineur.
- 2) Reportez-vous à la table des coefficients de variation pour le CANADA.
- 3) Le numérateur de cette estimation de rapport est 14 728 940. Le chiffre qui s'en rapproche le plus est 15 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 1,5 %.
- 4) Le dénominateur de cette estimation de rapport est 6 676 981. Le chiffre qui s'en rapproche le plus est 7 000 000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette rangée, soit 3,3 %.
- 5) Le coefficient de variation approximatif de l'estimation du rapport est donc donné par la règle 4, qui est :

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement.

C'est-à-dire que :

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0,015)^2 + (0,033)^2} \\ &= \sqrt{0,000225 + 0,001089} \\ &= 0,036 \end{aligned}$$

- 6) Le rapport obtenu entre les personnes ayant eu besoin de services de santé de routine ou de suivi et celles ayant eu besoin de soins immédiats nécessaires pour un problème de santé mineur est 14 728 940 / 6 676 981, c'est-à-dire 2,21 (à être arrondi selon les lignes directrices pour l'arrondissement figurant à la section 9.1). Le coefficient de variation de cette estimation est 3,6 %, ce qui fait qu'on peut la diffuser sans réserve.

Exemple 5 : Estimations de différences de rapports

Supposons qu'un utilisateur estime que le rapport entre les personnes ayant eu besoin de services de santé de routine ou de suivi et celles ayant eu besoin de soins immédiats nécessaires pour un problème de santé mineur au cours des 12 derniers mois s'établit à 1,84 pour les gens de la Colombie-Britannique, tandis qu'il est de 2,41 pour ceux du Québec. L'utilisateur est intéressé à comparer les deux rapports pour voir s'il y a une différence statistique entre ceux-ci. Comment l'utilisateur détermine-t-il le coefficient de variation de cette différence?

- 1) Tout d'abord calculez le coefficient de variation approximatif pour le rapport de la Colombie-Britannique (\hat{R}_1) et le rapport du Québec (\hat{R}_2), tel qu'il est décrit dans l'exemple 4. Reportez-vous les tables des coefficients de variation pour la Colombie-Britannique et le Québec. Le CV approximatif pour le rapport de la Colombie-Britannique est 7,9 % et 7,4 % pour celui du Québec.

- 2) En utilisant la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{R}_1 - \hat{R}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{R}_1 \alpha_1)^2 + (\hat{R}_2 \alpha_2)^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{R}_1 et \hat{R}_2 respectivement. C'est-à-dire que l'erreur-type de la différence $\hat{d} = 1,84 - 2,41 = -0,57$ est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(1,84)(0,079)]^2 + [(2,41)(0,074)]^2} \\ &= \sqrt{(0,021130) + (0,031805)} \\ &= 0,230 \end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,230 / (-0,57) = -0,404$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 40,4 %. La différence entre les estimations est considérée inacceptable et Statistique Canada recommande de ne pas la publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre F (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les utilisateurs subséquents en garde contre les hauts taux d'erreur associés à l'estimation.

L'Enquête canadienne sur l'expérience des soins de santé primaires, 2007-2008 - Fichier partagé

Tables de variabilité d'échantillonnage approximative - Québec

NUMÉRATEUR DU PERCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	307,3	305,9	304,4	299,7	291,7	283,5	275,0	266,3	257,2	247,9	238,2	217,4	168,4	97,2
2	217,3	216,3	215,2	211,9	206,3	200,4	194,5	188,3	181,9	175,3	168,4	153,7	119,1	68,8
3	177,4	176,6	175,7	173,0	168,4	163,7	158,8	153,7	148,5	143,1	137,5	125,5	97,2	56,1
4	153,7	153,0	152,2	149,8	145,8	141,7	137,5	133,1	128,6	123,9	119,1	108,7	84,2	48,6
5	137,4	136,8	136,1	134,0	130,4	126,8	123,0	119,1	115,0	110,9	106,5	97,2	75,3	43,5
6	125,5	124,9	124,3	122,3	119,1	115,7	112,3	108,7	105,0	101,2	97,2	88,8	68,8	39,7
7	****	115,6	115,0	113,3	110,2	107,1	103,9	100,6	97,2	93,7	90,0	82,2	63,7	36,7
8	****	108,2	107,6	106,0	103,1	100,2	97,2	94,1	91,0	87,6	84,2	76,9	59,5	34,4
9	****	102,0	101,5	99,9	97,2	94,5	91,7	88,8	85,7	82,6	79,4	72,5	56,1	32,4
10	****	96,7	96,3	94,8	92,2	89,6	87,0	84,2	81,3	78,4	75,3	68,8	53,3	30,7
11	*****	92,2	91,8	90,4	87,9	85,5	82,9	80,3	77,6	74,7	71,8	65,6	50,8	29,3
...
...
...
300	****	****	****	17,3	16,8	16,4	15,9	15,4	14,9	14,3	13,8	12,6	9,7	5,6
350	****	****	****	****	15,6	15,2	14,7	14,2	13,8	13,3	12,7	11,6	9,0	5,2
400	****	****	****	****	14,6	14,2	13,8	13,3	12,9	12,4	11,9	10,9	8,4	4,9
450	****	****	****	****	13,8	13,4	13,0	12,6	12,1	11,7	11,2	10,2	7,9	4,6
500	****	****	****	****	13,0	12,7	12,3	11,9	11,5	11,1	10,7	9,7	7,5	4,3
750	****	****	****	****	****	10,4	10,0	9,7	9,4	9,1	8,7	7,9	6,1	3,6
1,000	****	****	****	****	****	****	8,7	8,4	8,1	7,8	7,5	6,9	5,3	3,1
1,500	****	****	****	****	****	****	****	6,9	6,6	6,4	6,1	5,6	4,3	2,5
2,000	****	****	****	****	****	****	****	****	****	5,5	5,3	4,9	3,8	2,2
3,000	****	****	****	****	****	****	****	****	****	****	****	4,0	3,1	1,8
4,000	****	****	****	****	****	****	****	****	****	****	****	****	****	1,5
5,000	****	****	****	****	****	****	****	****	****	****	****	****	****	1,4

Note : Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

L'Enquête canadienne sur l'expérience des soins de santé primaires, 2007-2008 - Fichier partagé

Tables de variabilité d'échantillonnage approximative - Colombie-Britannique

NUMÉRATEUR DU PERCENTAGE (en milliers)	POURCENTAGE ESTIMÉ													
	0,1%	1,0%	2,0%	5,0%	10,0%	15,0%	20,0%	25,0%	30,0%	35,0%	40,0%	50,0%	70,0%	90,0%
1	247,0	245,9	244,6	240,9	234,4	227,8	221,0	214,0	206,8	199,2	191,4	174,7	135,4	78,1
2	174,7	173,9	173,0	170,3	165,8	161,1	156,3	151,3	146,2	140,9	135,4	123,6	95,7	55,3
3	142,6	142,0	141,2	139,1	135,4	131,5	127,6	123,6	119,4	115,0	110,5	100,9	78,1	45,1
4	****	122,9	122,3	120,4	117,2	113,9	110,5	107,0	103,4	99,6	95,7	87,4	67,7	39,1
5	****	110,0	109,4	107,7	104,8	101,9	98,8	95,7	92,5	89,1	85,6	78,1	60,5	34,9
6	****	100,4	99,9	98,3	95,7	93,0	90,2	87,4	84,4	81,3	78,1	71,3	55,3	31,9
7	****	92,9	92,5	91,0	88,6	86,1	83,5	80,9	78,1	75,3	72,3	66,0	51,2	29,5
8	****	86,9	86,5	85,2	82,9	80,6	78,1	75,7	73,1	70,4	67,7	61,8	47,9	27,6
9	****	82,0	81,5	80,3	78,1	75,9	73,7	71,3	68,9	66,4	63,8	58,2	45,1	26,0
10	****	77,8	77,4	76,2	74,1	72,0	69,9	67,7	65,4	63,0	60,5	55,3	42,8	24,7
11	****	74,1	73,8	72,6	70,7	68,7	66,6	64,5	62,3	60,1	57,7	52,7	40,8	23,6
...
...
...
300	****	****	****	****	13,5	13,2	12,8	12,4	11,9	11,5	11,1	10,1	7,8	4,5
350	****	****	****	****	****	12,2	11,8	11,4	11,1	10,6	10,2	9,3	7,2	4,2
400	****	****	****	****	****	11,4	11,1	10,7	10,3	10,0	9,6	8,7	6,8	3,9
450	****	****	****	****	****	10,7	10,4	10,1	9,7	9,4	9,0	8,2	6,4	3,7
500	****	****	****	****	****	10,2	9,9	9,6	9,2	8,9	8,6	7,8	6,1	3,5
750	****	****	****	****	****	****	****	7,8	7,5	7,3	7,0	6,4	4,9	2,9
1,000	****	****	****	****	****	****	****	****	6,5	6,3	6,1	5,5	4,3	2,5
1,500	****	****	****	****	****	****	****	****	****	****	****	4,5	3,5	2,0
2,000	****	****	****	****	****	****	****	****	****	****	****	****	3,0	1,7
3,000	****	****	****	****	****	****	****	****	****	****	****	****	****	1,4

Note : Pour bien utiliser ces tables, veuillez consulter la documentation sur les microdonnées.

10.2 Comment utiliser les tables de coefficients de variation pour obtenir des limites de confiance

Bien que les coefficients de variation soient beaucoup utilisés, l'intervalle de confiance d'une estimation est une mesure plus intuitivement significative de l'erreur d'échantillonnage. Un intervalle de confiance constitue une déclaration du niveau de confiance selon laquelle la valeur vraie pour la population se situe à l'intérieur d'une gamme précisée de valeurs. Par exemple, un intervalle de confiance de 95 % peut être décrit comme suit :

Si l'échantillonnage de la population est répété indéfiniment, chaque échantillon menant à un nouvel intervalle de confiance pour une estimation, l'intervalle englobera alors dans 95 % des échantillons la valeur vraie de la population.

En utilisant l'erreur-type d'une estimation, des intervalles de confiance pour des estimations peuvent être obtenues en partant de l'hypothèse qu'aux termes d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population se répartiront normalement autour de la valeur vraie de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que l'écart entre une estimation de l'échantillon et la valeur vraie pour la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types et environ 99 chances sur 100 que l'écart soit inférieur à trois erreurs-types. Ces différents degrés de confiance sont désignés sous le nom de niveaux de confiance.

Des intervalles de confiance pour une estimation \hat{X} sont généralement exprimés sous forme de deux chiffres, un inférieur et un supérieur à l'estimation, comme étant $(\hat{X} - k, \hat{X} + k)$, où k est déterminé suivant le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

Des intervalles de confiance pour une estimation peuvent être calculés directement à partir des tables de variabilité d'échantillonnage approximative, en déterminant d'abord à partir de la table appropriée le coefficient de variation de l'estimation \hat{X} , puis en utilisant la formule suivante pour le convertir à un intervalle de confiance ($IC_{\hat{x}}$) :

$$IC_{\hat{x}} = (\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}})$$

où $\alpha_{\hat{x}}$ est le coefficient de variation déterminé de \hat{X} , et

- $t = 1$ si l'on désire un intervalle de confiance de 68 %;
- $t = 1,6$ si l'on désire un intervalle de confiance de 90 %;
- $t = 2$ si l'on désire un intervalle de confiance de 95 %;
- $t = 2,6$ si l'on désire un intervalle de confiance de 99 %.

Nota : Les lignes directrices pour la diffusion des estimations s'appliquent également aux intervalles de confiance. S'il est impossible, par exemple, de diffuser une estimation, on ne peut alors pas non plus communiquer un intervalle de confiance.

10.2.1 Exemple d'utilisation des tables de coefficients de variation pour obtenir des limites de confiance

Un intervalle de confiance de 95 % pour la proportion estimée de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois qui déclareraient avoir eu de la difficulté à obtenir ces services (d'après l'exemple 2 à la section 10.1.1) serait calculé comme suit :

$$\hat{X} = 13,1 \% \text{ (ou exprimé sous forme de proportion } 0,131)$$

$$t = 2$$

$$\alpha_{\hat{x}} = 6,8 \% \text{ (0,068 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, tel que déterminé à partir des tables.}$$

$$CI_{\hat{x}} = \{0,131 - (2) (0,131) (0,068), 0,131 + (2) (0,131) (0,068)\}$$

$$CI_{\hat{x}} = \{0,131 - 0,018, 0,131 + 0,018\}$$

$$CI_{\hat{x}} = \{0,113, 0,149\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 11,3 % et 14,9 % des personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois avaient eu de la difficulté à obtenir ces services.

10.3 Comment utiliser les tables de coefficients de variation pour effectuer un test t

Des erreurs-types peuvent aussi être utilisées pour effectuer des tests d'hypothèses, une procédure destinée à distinguer des paramètres d'une population à l'aide d'estimations d'un échantillon. Ces estimations peuvent être des chiffres, des moyennes, des pourcentages, des rapports, etc. Les tests peuvent être effectués à divers niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Supposons que \hat{X}_1 et \hat{X}_2 sont des estimations d'un échantillon pour deux caractéristiques qui nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ est $\sigma_{\hat{d}}$.

Si $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ se situe entre -2 et 2, aucune conclusion à propos de la différence entre les

caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

10.3.1 Exemple d'utilisation des tables de coefficients de variation pour effectuer un test t

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de personnes ayant eu besoin de services de santé de routine ou de suivi au cours des 12 derniers mois qui déclaraient avoir eu de la difficulté à obtenir ces services chez les personnes ayant déclaré avoir un médecin habituel et chez les autres personnes. D'après l'exemple 3 à la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,051. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0,125 - 0,250}{0,051} = \frac{-0,125}{0,051} = -2,45$$

Puisque $t = -2,45$ est inférieur à -2 , il faut en conclure qu'il existe une différence significative entre les deux estimations au niveau de signification de 0,05.

10.4 Coefficients de variation pour des estimations quantitatives

Il faudrait produire des tables spéciales afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que la plupart des variables pour l'ECE-SSP 2007-2008 sont principalement de nature catégorique.

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Par exemple, le coefficient de variation du nombre total de fois où les gens ont personnellement utilisé le service d'urgence d'un hôpital au cours des 12 derniers mois serait supérieur au coefficient de variation de la proportion correspondante de personnes les ayant utilisées. Si, par conséquent, le coefficient de variation de la proportion est inacceptable (rendant la proportion non diffusable), il en sera de même du coefficient de variation de l'estimation quantitative correspondante (rendant cette estimation quantitative non diffusable).

Des coefficients de variation de telles estimations peuvent être calculés, au besoin, pour une estimation précise à l'aide d'une technique appelée pseudo-répétition, ce qui veut dire diviser les enregistrements inclus dans les fichiers de microdonnées en sous-groupes (ou répétitions) et déterminer la variation à l'intérieur de l'estimation de répétition en répétition. Les utilisateurs qui désirent calculer des coefficients de variation pour des estimations quantitatives peuvent communiquer avec Statistique Canada afin d'obtenir des conseils sur l'allocation d'enregistrements à des répétitions appropriées et sur les formules à employer à l'intérieur de ces calculs.

10.5 Tables des coefficients de variation

Consulter le fichier ECE-SSP2007-2008_CVTabF.pdf pour les tables de coefficients de variation.

10.6 Méthode bootstrap pour estimer la variance

Pour juger de la qualité d'une estimation et en calculer le CV, on doit d'abord établir l'écart-type. On a aussi besoin de l'écart-type d'estimations pour les intervalles de confiance. Dans l'ECE-SSP, on utilise des plans d'échantillonnage et d'étalonnage à plusieurs degrés et, par conséquent, il n'y a pas de formule simple d'estimation de variance qu'on puisse appliquer. Il faut donc procéder par approximation. Si on se sert de la méthode bootstrap, c'est qu'il faut tenir compte de l'échantillonnage et de l'étalonnage au moment d'estimer la variance. C'est ce que fait la méthode bootstrap et, avec le programme Bootvar dont il sera question à la sous-section suivante, l'utilisateur dispose déjà d'une méthode d'une application assez facile.

Dans l'ECE-SSP, on emploie la méthode bootstrap que décrit W. Yung (Yung, W. (1997b) Estimation de la variance des fichiers de microdonnées à grande diffusion, *Symposium 1997, Nouvelles orientations pour les enquêtes et les recensements*, Statistique Canada).

On prélève indépendamment sur chaque strate un échantillon aléatoire simple de $(n - 1)$ des unités n de l'échantillon. À noter que, comme la sélection s'opère avec remise, une unité peut être choisie plusieurs fois. Toute l'opération (prélèvement d'échantillons aléatoires simples et repondération de chaque strate) a lieu à B reprises, B étant une valeur élevée, ce qui donne B valeurs initiales de pondération bootstrap. L'ECE-SSP utilise $B = 500$ pour produire 500 valeurs de pondération.

On repondère alors comme dans la pondération courante par correction de non-réponse, étalonnage, etc. On obtient 500 valeurs de pondération bootstrap finale pour chaque unité de l'échantillon. On rapporte la variation des 500 estimations possibles correspondant aux 500 valeurs bootstrap à la variance de l'estimateur par pondération courante; la variance peut être estimée par ce moyen. Pour plusieurs raisons, l'utilisateur pourrait vouloir calculer le CV des estimations par la méthode bootstrap. En voici quelques-unes :

- Premièrement, s'il désire des estimations à un niveau géographique inférieur à celui de la province (au niveau des régions urbaines et rurales, par exemple), les Tables de la variabilité d'échantillonnage approximative ne peuvent suffire. Par le programme d'estimation de variance bootstrap, on peut recourir aux techniques d'estimation de domaine pour dégager le CV de ces estimations.
- Deuxièmement, si l'utilisateur veut une analyse plus fine par régression linéaire ou logistique pour l'estimation des coefficients, les Tables de la variabilité d'échantillonnage approximative ne lui donneront pas les CV correspondants en toute précision. Bien qu'un certain nombre de progiciels statistiques permettent d'intégrer la pondération d'échantillonnage à l'analyse, les variances produites ne tiennent souvent pas tout à fait compte du plan d'échantillonnage ni ne traduisent l'étalonnage de pondération contrairement à ce qui se passe dans le programme d'estimation de variance bootstrap.
- Troisièmement, dans le cas des estimations de variables quantitatives, il faut consulter des tableaux séparés pour établir l'erreur d'échantillonnage.

10.7 Progiciels statistiques pour estimer la variance

Statistique Canada a élaboré un programme qui peut livrer des estimations de variance bootstrap. C'est le programme Bootvar.

Celui-ci est disponible en format SAS. Il est formé de macro instructions d'estimation de variance pour les totaux, les rapports et les différences entre rapports, ainsi que de régression linéaire et logistique.

Le Bootvar peut être téléchargé à partir du site Internet des Centres de données de recherche (CDR) de Statistique Canada. Il faut accepter la licence d'adhésion automatique Bootvar avant de pouvoir lire les fichiers.

SAS: http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-fra.htm

SPSS: http://www.statcan.gc.ca/rdc-cdr/bootvar_spss-fra.htm

10.7.1 Autres progiciels

Une variable du poids d'enquête ayant une série correspondante de 500 variables de poids bootstrap accompagne de nombreux fichiers de données de l'ECE-SSP afin qu'une approche entièrement fondée sur le plan d'enquête puisse être adoptée pour l'analyse des données.

L'approche de l'analyse fondée sur le plan d'enquête suppose d'abord l'utilisation de la variable du poids d'enquête en vue d'obtenir les estimations pondérées des quantités d'intérêt. On utilise ensuite des renseignements additionnels sur le plan d'enquête pour estimer les variances (et covariances) des quantités estimées. Dans le cas des fichiers de microdonnées à grande diffusion (FMGD) de l'ECE-SSP, ces renseignements additionnels comprennent 500 variables de poids bootstrap de l'enquête. On peut alors utiliser les estimations et estimations des variances fondées sur le plan d'enquête pour tirer les conclusions nécessaires dans le cadre de l'analyse.

On peut décrire brièvement la création d'une estimation de variance de poids bootstrap de la façon suivante :

Supposons que $\hat{\beta}$ soit l'estimation pondérée de la quantité d'intérêt, β , calculée au moyen de la variable du poids de l'enquête w , et que $\hat{\beta}^{(b)}$ soit une estimation obtenue exactement de la même façon, à l'exception de la substitution b^{th} de la variable de poids bootstrap $w^{(b)}$ pour la variable de poids d'enquête w , $b=1,2,\dots,500$. Cela donne des estimations bootstrap w de $\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(500)}$. L'estimation du poids bootstrap de la variance de $\hat{\beta}$ est donc

$$\hat{V}_B(\hat{\beta}) = \frac{1}{500} \sum_{b=1}^{500} (\hat{\beta}^{(b)} - \hat{\beta})^2 \quad (1)$$

Si $\hat{\beta}$ est un vecteur plutôt qu'une valeur unique, comme si $\hat{\beta}$ était l'ensemble des coefficients d'un modèle, alors la matrice des estimations des variances et covariances des éléments de $\hat{\beta}$ est $\hat{V}_B(\hat{\beta}) = \frac{1}{500} \sum_{b=1}^{500} (\hat{\beta}^{(b)} - \hat{\beta})(\hat{\beta}^{(b)} - \hat{\beta})'$. (La valeur « 500 » figurant dans la formule provient du fait que nous avons 500 poids bootstrap moyens différents.

La méthode bootstrap n'est qu'une méthode de répétition parmi d'autres que l'on peut utiliser pour obtenir des estimations de variances fondées sur le plan d'enquête à partir des données d'enquête. Dans les sections ci-dessous, des directives seront données sur la mise en oeuvre de l'estimation de la variance des poids bootstrap à partir des données des FMGD de l'ECE-SSP, au moyen de trois progiciels commerciaux différents qui peuvent effectuer une certaine analyse fondée sur le plan d'enquête pour le BRR:

- Stata 9 ou 10,
- SUDAAN et
- WesVar.

Ces méthodes sont adaptées pour l'ECE-SSP d'après un article d'Owen Phillips intitulé *Comment utiliser les poids bootstrap avec Wes Var et SUDAAN* (n° 12-002-X20040027032 au catalogue) paru dans « *Le Bulletin technique et d'information des Centres de données de recherche (Index chronologique)* », automne 2004, vol.1 no 2, Statistique Canada, n° 12-002-XIF au catalogue. Dans le fichier de données de l'ECE-SSP où des poids bootstrap sont fournis, les noms donnés à ces variables bootstrap dans la documentation de l'utilisateur sont de **wrps0001** à **wrps0500**. Le nom de la variable du poids d'enquête est soit **wtps**.

Stata 9 ou 10

À partir de la version 9, le progiciel commercial Stata contient des méthodes de répétition additionnelles permettant d'effectuer des estimations de variances fondées sur le plan d'enquête dans ses commandes d'analyse d'enquêtes. La méthode BRR fait partie des méthodes de répétition offertes; c'est cette méthode qui serait précisée dans l'analyse des données de l'ECE-SSP. Pour préciser cette méthode, voici la marche à suivre recommandée :

1. Avant d'utiliser les commandes d'analyse d'enquêtes, utiliser l'énoncé « **svyset** » pour déclarer que les données sont des données d'enquête, pour désigner les variables qui contiennent des renseignements sur le plan d'enquête et pour préciser la méthode d'estimation des variances. Les paramètres établis par l'énoncé « **svyset** » sont sauvegardés avec un ensemble de données lorsque (ou si) un ensemble de données est sauvegardé. La forme de l'énoncé « **svyset** » à utiliser avec l'ensemble de données d'une analyse de l'ECE-SSP serait la suivante :

svyset [pweight=wtps], vce(brr) brrweight(wrps0001-wrps0500) mse

L'énoncé **pweight=wtps** indique à Stata que le poids d'enquête (souvent appelé le poids de probabilité) est la variable **wtps**. L'option **vce(brr)** déclare que la méthode d'estimation de la variance à utiliser est BRR. L'option **brrweight(wrps0001-wrps0500)** indique que le nom des variables du poids BRR sont **wrps0001**, **wrps0002**, ..., **wrps0500**. Cette option peut aussi être appelée **brrweight(wrps0*)** à condition qu'il n'y ait pas de variable autre que les variables du poids bootstrap dont le nom commence par « wrps0 ».

Enfin, l'option **mse** indique à Stata de calculer la variance au moyen des différences au carré entre les estimations bootstrap et l'estimation de l'échantillon complet des variables, comme le montre l'équation (1). Si cette option n'est pas incluse, Stata utilise les différences au carré entre chaque estimation bootstrap et la moyenne de toutes les estimations bootstrap. Les deux méthodes devraient produire à peu près le même résultat.

2. Stata contient une liste complète de commandes d'analyse d'enquêtes qui comporte une méthode fondée sur le plan d'enquête pour ses calculs. Ces commandes, décrites dans les documents Stata, sont mises en oeuvre par l'utilisation du préfixe « **svy** » associé au nom des autres estimateurs. Par exemple, **svy: mean** est la commande servant à estimer les moyennes et les estimations de variabilité de la population et des sous-populations grâce à une méthode fondée sur le plan d'enquête. Lorsque l'énoncé **svyset** précède toutes les commandes de l'enquête, il n'est pas nécessaire que celles-ci contiennent des renseignements sur la méthode fondée sur le plan d'enquête à adopter. Il convient de noter que la

plupart des commandes qui permettent le préfixe « svy » correspondent également au nom des commandes relatives aux données autres que les données d'enquête, à ce qui est estimé, aux options offertes et à ce qui peut être fait par le changement apporté après l'estimation lorsque le préfixe « svy » est ajouté.

SUDAAN

SUDAAN est un progiciel commercial expressément élaboré par le Research Triangle Institute pour l'analyse de données tirées d'enquêtes-échantillons complexes et d'autres études par observation et expérimentales comprenant des données corrélées par grappes. La version exécutable SAS du logiciel est particulièrement utile aux personnes qui connaissent le SAS. Dans la version 9.0 et les versions ultérieures, toutes les procédures indiquées dans SUDAAN peuvent utiliser la méthode BRR pour estimer des variances et des covariances.

La spécification de la méthode d'estimation des variances à utiliser par SUDAAN est apportée dans la déclaration de procédure pour une procédure particulière. D'autres déclarations sur les plans d'échantillonnage fournissent des renseignements supplémentaires requis par le programme. En particulier, pour effectuer un amorçage à partir des données de l'ECE-SSP, il faut suivre les étapes suivantes :

- préciser **DESIGN=BRR** dans la déclaration de procédure;
- inclure l'énoncé du POIDS suivant pour définir la variable du poids d'enquête : **WEIGHT wtps**;
- inclure l'énoncé REPWGT pour indiquer le nom des variables de poids bootstrap sur le fichier de données. Plus particulièrement, pour les FMGD de l'ECE-SSP, cet énoncé REPWGT aurait la forme suivante :

REPWGT wrps0001-wrps0500;

WesVar

WesVar est un progiciel produit par Westat qui effectue diverses analyses de données d'enquête en se fondant exclusivement sur des méthodes de répétition pour l'estimation des variances. L'une des méthodes offertes est la méthode BRR. Largement tirée de Phillips (2004), dans WesVar, la méthode d'estimation des variances est mentionnée lors de la création d'un nouveau fichier de données WesVar.

Le fichier qui en résulte sert à définir des classeurs lorsque l'on procède à des demandes de tableaux et de régressions. Pour définir un fichier de données WesVar avec un poids bootstrap :

- déplacer les variables de poids répétées (c.-à-d. wrps0001 à wrps0500) dans la boîte *Replicates (Répétitions)*;
- déplacer la variable du poids d'enquête (c.-à-d. wtps) dans la boîte *Full sample (Échantillon complet)*;
- déplacer les variables de l'analyse dans la boîte *Variables (Variables)*, un identificateur unique de la boîte d'identité (facultatif) et sauvegarder le fichier.

Phillips (2004) illustre ces directives par un exemple utilisant les données de cycle 14 de l'Enquête Sociale Générale.

11.0 Pondération

Puisque l'Enquête canadienne sur l'expérience des soins de santé primaires (ECE-SSP) 2007-2008 faisait appel à un sous-échantillon de l'échantillon de l'Enquête sur la santé dans les collectivités canadiennes (ESCC), cycle 4.1, le calcul des poids pour les enregistrements des données de l'enquête est clairement lié à la procédure de pondération utilisée aux fins de l'ESCC. La procédure de pondération employée pour l'ESCC est décrite brièvement ci-dessous.

11.1 Procédures de pondération pour l'Enquête sur la santé dans les collectivités canadiennes

Des bases de sondage aréolaire et téléphonique ont été utilisées pour l'ESCC 4.1. Dans l'ESCC, les répondants de chacune des deux bases sont pondérés séparément avant la combinaison des deux bases, et un rajustement est fait à des fins d'intégration. Le poids initial de l'ECE-SSP est celui des répondants sélectionnés pour l'ESCC, tel qu'il est calculé après la combinaison des bases, après la « winsorization » et juste avant la post-stratification. Ce poids est réputé représenter de façon appropriée l'ensemble de la population cible de l'enquête. La stratégie de pondération des unités de la base aréolaire de l'ESCC est détaillée dans le Guide de l'utilisateur du fichier de microdonnées à grande diffusion de l'ESCC 4.1. Le poids final intégré, avant la post-stratification, de la base aréolaire de l'ESCC 4.1 tient compte de la probabilité de sélection de chaque ménage, de la non-réponse au niveau des ménages, de la sélection de la personne choisie au sein du ménage et de la non-réponse au niveau des personnes.

11.2 Procédures de pondération pour l'Enquête canadienne sur l'expérience des soins de santé primaires

Le poids initial de l'ECE-SSP de 2007-2008 est le poids final intégré avant la post-stratification de l'ESCC 4.1. Ce poids est rajusté pour tenir compte de la sélection d'un échantillon restreint de répondants de l'ESCC, d'unités non déterminées et de la non-réponse à l'ECE-SSP 2007-2008. Les poids sont aussi ajustés pour contrer la présence de poids aberrants et pour s'assurer que les estimations de l'ECE-SSP 2007-2008 concordent avec les projections démographiques pour certains sous-groupes de la population. Tous ces ajustements sont expliqués dans la présente section.

Sélection de l'échantillon

Le poids initial, celui de l'ESCC 4.1, fournit une représentation adéquate de la population ciblée à condition d'inclure tous les répondants. Pour l'ECE-SSP 2007-2008, un échantillon de 16 482 répondants a été choisi aléatoirement parmi tous les répondants admissibles de la base aréolaire. L'échantillon a été choisi de façon indépendante pour chaque province et territoire à l'aide de l'échantillonnage aléatoire systématique. Ainsi, tous les répondants de l'ESCC âgés de 18 ans et plus d'une même province ou d'un même territoire avaient la même probabilité d'être choisis. Le poids de sélection de l'ECE-SSP 2007-2008 a été combiné au poids initial fourni par l'ESCC de façon à s'assurer que la somme des poids de tous les répondants à l'intérieur de chaque province/territoire soit préservée. On a eu recours à une méthode de partage des poids afin de générer les poids initiaux pour un très petit nombre d'unités d'appui échantillonnées du Québec.

Un petit échantillon du Nord a été ajouté afin que cette région contribue à l'estimation nationale, mais il faut noter que les estimations ne devraient pas être produites à ce niveau. Même la combinaison des trois territoires générerait probablement des estimations qui ne sont pas publiables pour l'ensemble du Nord.

Ainsi, le poids de sélection rajusté de l'ECE-SSP 2007-2008 est donné par :

$$\text{Poids de l'ECE – SSP} = \left(\frac{\text{Poids de l'ESCC 4.1} * \sum \text{poids de tous les répondants de l'ESCC 4.1 âgés de 18 ans et plus dans la prov}}{\sum \text{poids de tous les répondants sélectionnés pour l'ECE – SSP dans la province}} \right)$$

Rajustement pour les cas déterminés

Pour diverses raisons, certaines personnes n'ont pu être interviewées pour l'ECE-SSP de 2007-2008. Dans certains cas, des coordonnées à jour n'étaient pas disponibles. Dans d'autres cas, la période de collecte s'est terminée sans que le répondant ait pu être joint. En outre, d'autres personnes ont refusé de participer à l'enquête. Une partie de l'échantillon initial de l'ECE-SSP est donc « perdu », et des facteurs de rajustement ont dû être appliqués aux poids des répondants pour compenser cette non-réponse ou cette absence de contact.

Le premier rajustement a été réalisé pour les cas indéterminés au moment de l'Enquête (c.-à-d. les cas pour lesquels on n'a pu déterminer s'ils faisaient dans le champ ou hors du champ de l'enquête parce qu'on n'a pu les contacter). Les poids des unités indéterminées sont redistribués aux unités déterminées au sein des groupes d'unités déterminées. Pour ce faire, on a recours à la régression logistique. Un modèle de prédiction de la probabilité de détermination des unités a été bâti à l'aide des variables disponibles pour toutes les personnes choisies à l'ECE-SSP. Étant donné la grande quantité d'information disponible de l'ESCC, un vaste éventail de possibilités s'offrait pour bâtir le modèle de détermination des unités. À l'aide de ce modèle, les répondants ont été divisés en 12 groupes, selon leur probabilité de détermination dans l'enquête. Des groupes comprenant un nombre égal d'unités déterminées ont été créés. Chaque unité indéterminée a ensuite été ajoutée au groupe qui correspondait à sa propre probabilité de détermination. Dans chaque groupe, le poids de l'unité déterminée était ensuite augmenté d'un facteur correspondant à la somme des poids de toutes les unités dans le groupe d'unités déterminées, divisé par le poids de toutes les unités dans le groupe.

Rajustement pour la non-réponse

Étant donné que le taux de réponse était très élevé pour l'enquête, il n'y avait pas beaucoup d'unités disponibles pour établir un modèle de régression pour la non-réponse. Un modèle simple fondé sur la région (les provinces de l'Atlantique, le Québec, l'Ontario, l'Alberta, la Colombie-Britannique et le Yukon, et ainsi que les autres provinces), sur l'âge (18 à 34 ans, 35 à 49 ans, 50 à 64 ans et 65 ans et plus) et le sexe a été adopté pour tenir compte de la non-réponse au sein de ces groupes. Le modèle pour les autres territoires était fondé uniquement sur le sexe. Ce rajustement tenait également compte des répondants qui n'ont pas consenti à ce que l'on communique et/ou lie leurs données.

Contrôle des poids aberrants

Étant donné la série d'ajustements successifs appliqués sur les poids des répondants, d'abord par l'ESCC puis par l'ECE-SSP 2007-2008, il est possible que certaines unités se retrouvent avec des poids se démarquant des poids des autres répondants du même groupe démographique au point même de devenir aberrants. Certains répondants peuvent effectivement représenter une proportion anormalement élevée de leur groupe et ainsi influencer fortement les estimations pour ces groupes. Afin d'éviter cette situation, le poids des répondants qui contribuent de façon aberrante à leur groupe démographique est ajusté à la baisse selon une méthode dite de « winsorization ».

Les groupes fondés sur l'âge (18 à 34 ans, 35 à 49 ans, 50 à 64 ans et 65 ans et plus) et le sexe, combinés à

- a. la région (les provinces de l'Atlantique, le Québec, l'Ontario, l'Alberta, la Colombie-Britannique et le Yukon ainsi que les autres provinces) ou
 - b. la province,
- ont été examinés.

Très peu d'unités ont vu leur poids être ainsi modifié.

Post-stratification

La dernière étape nécessaire afin d'obtenir le poids final de l'ECE-SSP est la post-stratification. La post-stratification est appliquée afin d'assurer que la somme des poids finaux corresponde aux estimations de population pour chacun des 82 groupes par province (trois territoires combinés) et des quatre groupes d'âge (18 à 24 ans, 25 à 44 ans, 45 à 64 ans, 65 ans et plus), pour un seul groupe d'âge dans les territoires (18 ans et plus) et pour chaque sexe. Les estimations de population en date du 17 mai 2008 ont été utilisées pour la post-stratification. Le poids final de l'ECE-SSP 2007-2008 est donné par :

$$\text{Poids final pour ECE - SSP} = \left(\frac{\text{Poids winsorié} * \text{projection démographique pour le groupe auquel appartient le répondant}}{\sum \text{poids de tous les répondants dans le groupe}} \right)$$

Le poids qui en résulte WTPS (pour « final weight » en anglais) est le poids final qui figure dans le fichier partagé de microdonnées de l'ECE-SSP 2007-2008.

Le poids qui en résulte WTPP (pour « final weight » en anglais) est le poids final qui figure dans le fichier de microdonnées à grande diffusion de l'ECE-SSP 2007-2008.

12.0 Questionnaire

Consulter le fichier ECE-SSP2007-2008_QuestF.pdf pour la version française du questionnaire utilisé pour la collecte des données.

13.0 Cliché d'enregistrement à valeurs univariées

Consulter le fichier ECE-SSP2007-2008_LvCds.pdf pour le cliché d'enregistrement à chiffres univariés.